TABLE OF CONTENTS

Section

Page

TABLE OF FIGURES

Exhibit
Page

1. EXECUTIVE SUMMARY

This technical analysis report (TAR) presents the results of an independent assessment, conducted by the EOSDIS IV&V team during the period from 15 September 1994 to 31 January 1995, to evaluate the modeling activity being performed by the ECS development contractor, Hughes Applied Information Systems (HAIS). The objective of the independent analysis is to provide the ESDIS Project with objective insights into the validity of the underlying assumptions and predictive quality of the models to support the ECS Preliminary Design Review (PDR) evaluation by the Government. The results presented in this report focus on the state of the models (i.e., the level of maturity attained) at the time they were used to generate predictions in support of the PDR. Detailed analyses of the system-level modeling activities associated with user loading and demand (User Model), production loading (Production Model), system architecture and performance (Performance Model), and cost estimation (Cost Model) have been conducted. A summary of findings for each of these areas follows.

User Model

The User Model, also referred to as the "pull" model, describes the interaction of Earth science researchers and other users with the EOSDIS. It characterizes the user load by describing who the users are, what information they need and when, and how they are likely to interact with the EOSDIS. The IV&V analysis focused on the following aspects of User Modeling: user characterization, user scenarios, user services, and product access requirements.

The characterization of user demography and product demand is still in progress by the ECS contractor. Therefore, the results presented here are preliminary and based on a snapshot in time of ongoing User Modeling activities. The analysis was further constrained by delays in IV&V accessibility to the PDR Technical Baseline and availability of ECS contractor User Modeling documentation. To date, the HAIS characterization of user demography appears reasonable, however the maturity level of this activity is somewhat limited pending completion of an ongoing survey. The user scenarios do not adequately represent the expected distribution of science users; this could potentially result in incorrect estimates of user service and data requirements that are key inputs into the Performance Model. Although user services information (e.g., service types, frequency of access, anticipated users) identified thus far, adequately represents the scenarios, its usefulness is limited because it is based on scenarios that may not represent the interests of the entire user community. Accurate assessment of user requirements for data and products from the EOSDIS is an important input into the Performance Model. This activity has achieved limited maturity; information gathering and modeling in this area are ongoing. The IV&V analysis has identified a number of areas where additional work is needed to adequately represent the user characterization and requirements. Continued modeling efforts in this area are essential for achieving long-term user satisfaction.

Production Model

The Production Model, also referred to as the "push" model, describes the science generation process; nominally, the transformation of Level 0 input to Level 1, 2, 3, and 4 products. The purpose of the model is to predict the steady-state and exceptional processing necessary to deliver trusted science data to the EOSDIS archives when required. The IV&V analysis focused on a preliminary evaluation of the following: impact of the mission redefinition (following the System Design Review (SDR)) on the Production Model; instrument product availability at various epochs; data dependencies for product generation; and the translation of production information provided by the Ad Hoc Working Group on Production (AHWGP) to input parameters for the Performance Model.

Analysis results presented in this report are preliminary due to several constraints: the delayed availability of the model for the IV&V analysis; the version reviewed lacked functionality and did not incorporate all the data provided by the AHWGP; lack of documentation describing the ECS contractor's Production Modeling analyses; delayed accessibility to the PDR Technical Baseline; and the redefinition of ECS responsibilities resulting from the EOS Mission Profile rebaseline. Although there were several limitations, the analysis results clearly identify problem areas and potential issues. Areas of concern include the large discrepancies in the processing loads and data requirements provided by the AHWGP, the PDR Technical Baseline, and the resource allocation by the ESDIS Project. Issues relating to the immaturity of the model were also identified. For example, the version of the model analyzed does not fully account for all events that are expected to occur during the four PDR epochs. Additional analysis is recommended to further examine Production Modeling issues such as these as well as assess the technical integrity and user satisfaction aspects of the model.

Performance Model

The Performance Model is intended to provide a definitive basis for evaluating alternative ECS architectures capable of supporting user and production demands (as predicted by the User and Production Models), and evaluating architectural sensitivities to predictive uncertainties. This model is implemented using the Block Oriented Network Simulator (BONeS) discrete-event simulation modeling tool. The IV&V analysis focused on the evaluation and validation of the following: completeness and accuracy of the BONeS model representation of the system functions, distributed architecture, and push and "pull" workloads; derivation and usage of the model input parameters; soundness of performance statistics collection defined in the model; and the overall structure of the model to support evolvability.

Analysis results presented in this report are preliminary due to several factors: the delayed availability of the model for the IV&V analysis; lack of documentation describing the ECS contractor's Performance Modeling analyses; and the fact that no modeling results were delivered with the model. In light of these constraints, the analysis has identified problems and potential issues. The most immediate issue is the need for the ECS modeling team to enhance some areas and fully integrate the model so that an analysis of the push and "pull" workloads together can be performed. The consequence of not performing these enhancements is underestimation of delay and processor utilization, which could potentially result in underestimated costs. A second problem area is the need for a more accurate representation of system resources that are essential to accurately predict performance and resource sizing. In some cases the model representations will result in overestimated delays, and in other cases underestimated delays. Both of these cases have potential cost implications: overestimated or underestimated costs. Follow-on analysis is recommended to further examine Performance Modeling issues such as these using the methodology and evaluation criteria defined in this report.

Cost Model

The Cost Model provides resource estimates required to develop and operate ECS architectural alternatives (as partially derived from the Performance Model) within schedule constraints. The IV&V analysis focused on a preliminary evaluation of the Cost Model, implemented as a collection of three independent estimation models: Commercial Off-The-Shelf (COTS) hardware and software models, used to estimate hardware, software, and procurement costs, and perform cost impact analysis; a custom software model, used to estimate size, level of effort, schedules, and associated costs for custom developed software; and an operations and maintenance model, used to estimate personnel costs associated with ECS operations and maintenance.

Information to support the Cost Model analysis was obtained primarily through interviews and discussions with the ECS contractor. The analysis was constrained by not having the models and pertinent cost information in hand. Therefore, rather than draw firm conclusions, the results of the Cost Model analysis point to several areas where potential problems could arise. Key issues include: lack of a lifecycle Cost Model; trade studies that are based on restricted input; use of conservative parameters; and a custom software estimation approach that is untested. Given the current Cost Model implementation, there is no integrated mechanism to estimate lifecycle costs and subsequently perform what-if analyses. Without this capability, it is difficult to analyze the lifecycle benefits or drawbacks of implementing alternative solutions. The lack of an integrated lifecycle Cost Model could result in trade analyses based on only a subset of cost data, such as COTS hardware and software costs. The results can be misleading and actually lead to design decisions that increase, rather than decrease, costs. Of equal concern is the potential that some of the modeling parameters may be too conservative. Although some degree of conservatism may be warranted in the early stages of the lifecycle, overly conservative estimates can have negative implications such as the unnecessary de-scoping of systems. The final issue pertains to the custom software estimation approach. The method used is unproved and potentially underparameterized. Performing an independent second estimate using a different model is strongly recommended.

Conclusion

Our overriding conclusion is that the ECS User, Production, and Performance models, at the level of maturity attained to support PDR are too immature to provide reliable predictions of performance (see Exhibit 1-1). Material presented at PDR based upon these models (except for some very limited cases) is suspect. This should not be read as an indictment of the organizations or people who have strived to do an extremely difficult job in a very short period of time. Based upon our assessment of accomplishments to date, the effort could yield a viably mature set of user, production, and Performance Models by late-April to mid-May of this year.

Model Class	Model Element	PDR Maturity	Maturity Metrics
User Model	**Overall**	2	
	User Characterization (Who)	2	0: Nil Maturity
	User Access Rqts (What,When)	1	1: Limited Maturity
	User Scenarios (How)	2	2: Somewhat Limited
	User Services	2	3: Fully Mature
Production Model	**Overall**	2	
	TRMM: CERES	2?	D: Validated By The Developer(s)
	TRMM: LIS	2?	
	TRMM: VIRS, PR, TMI	2?	V: Validated By The IV&V Team
	AM-1: MODIS	2	
	AM-1: MISR	2	?: Estimated Based On Limited IV&V Inputs
	AM-1: MOPITT	2	
	AM-1: CERES	2	
	AM-1: ASTER	2	
	Landsat 7: ETM+	0?	
	Follow-On Missions: ADEOS,	0?	
	DAO (R. Rood's Products)	0?	
	V0: Migrated Datasets	0?	
Performance Model	**Overall**	1?	
	System Design Representation	1?	
	Workload Parameters	1?	
	Performance Statistics	1?	
	Model Structure	1?	
Cost Model	**Overall**	2?	
	COTS H/W & S/W	2?	
	Custom S/W	2?	
	Operations & Maintenance	2?	
	Model Interfaces (incl. Perf Mod)	1?	

EXHIBIT 1-1: Model Maturity Levels

To The Reader: If your planned reading of this document is limited to the Executive Summary, please consider also reading Section 2.4 (Background Information) to gain a full understanding of the context within which we arrived at our conclusions and recommendations.

2. INTRODUCTION

This section provides introductory information pertinent to this EOSDIS IV&V "EOSDIS Core System (ECS) Modeling Assessment Report". The purpose of the report, objectives and scope of the analysis, and relevant background information and references follow.

2.1 Purpose of the Report

The purpose of this technical analysis report (TAR) is to document the results of an independent assessment of the ECS contractor's modeling activity that was conducted by the EOSDIS IV&V team over the period 15 September 1994 to 31 January 1995. This TAR documents identified problems and potential issues, including their relative severity and possible adverse implications for employing the models to reliably predict ECS performance and cost (development and operation) estimates. This TAR is the second in a series of reports and follows the preliminary report [1] which was limited in scope to an independent EOSDIS user characterization assessment.

2.2 Objective of the Analysis

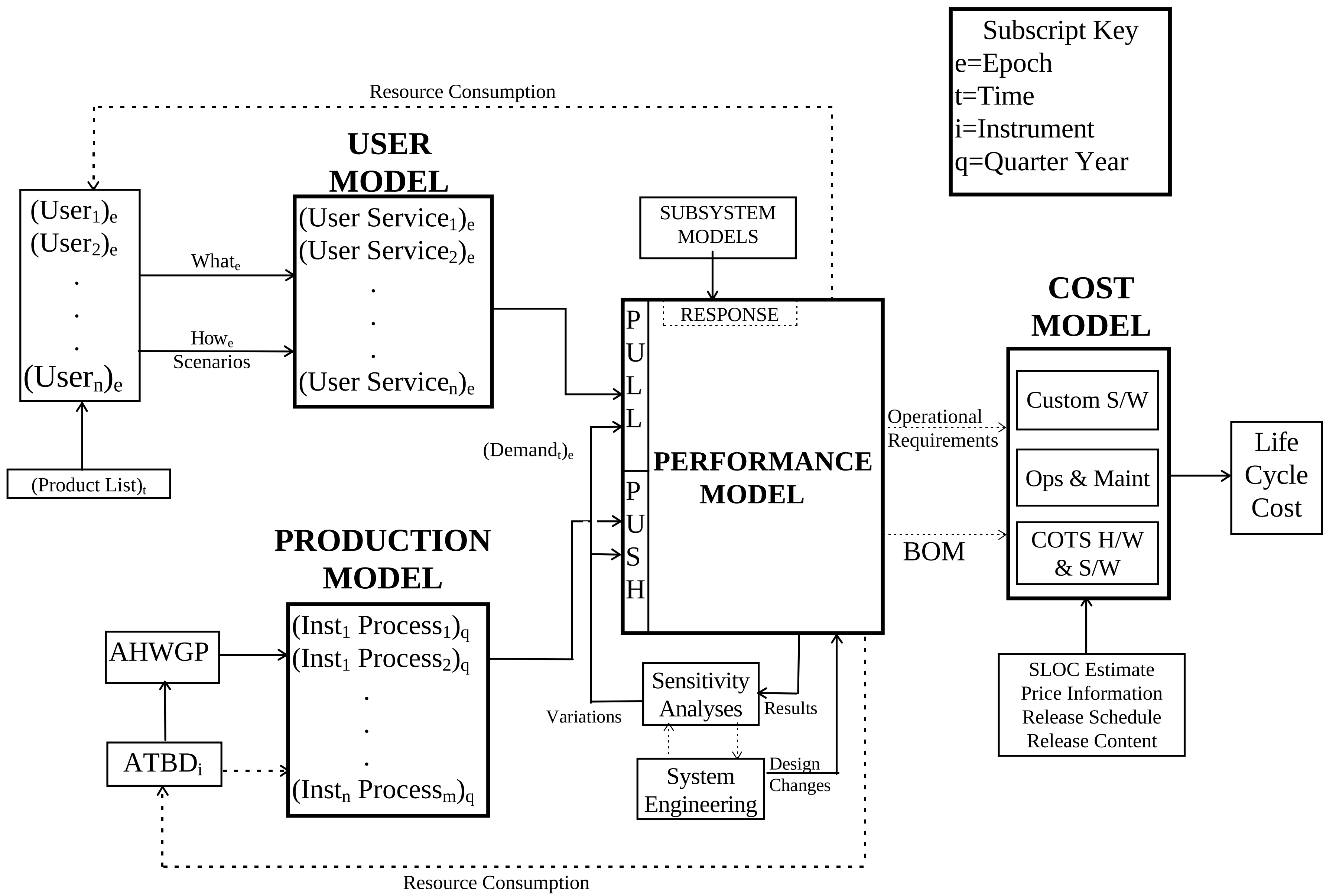
The objective of this analysis is to independently assess the ECS modeling activity being performed by the ECS development contractor (Hughes Applied Information Systems [HAIS]) to provide the ESDIS Project with objective insights into the validity of the underlying assumptions and predictive quality of the models to support the ECS Preliminary Design Review (PDR) evaluation by the Government.

2.3 Scope of the Analysis

This analysis examines system-level modeling activities associated with: 1) user and production loading, 2) system architecture and performance, and 3) cost estimation. The scope is limited to the state of the models (i.e., the level of maturity attained) at the time they were used to generate predictions in support of PDR. The analysis does not examine subsystem level modeling activities (i.e., DADS, etc.) which indirectly affect the quality of system level models by providing parametric drivers, typically in the form of subsystem response characteristics.

2.4 Background Information

The ECS modeling activity focuses on the formulation and implementation of four classes of interrelated models: the User Model, Production Model, Performance Model, and Cost Model. Exhibit 2-1 illustrates the context within which the models interact to generate predictions of performance and cost. The following discussion is keyed to this exhibit. The discussion does not address the correctness of the context. That is the subject of later sections of this TAR.



Subscript Key
e=Epoch
t=Time
i=Instrument
q=Quarter Year
Resource Consumption
USER MODEL
$(\text{User}_1)_e$
$(\text{User}_2)_e$
$(\text{User}_n)_e$
What_e
How_e Scenarios
$(\text{User Service}_1)_e$
$(\text{User Service}_2)_e$
$(\text{User Service}_n)_e$
$(\text{Product List})_t$
SUBSYSTEM MODELS
RESPONSE
PULL
PUSH
PERFORMANCE MODEL
$(\text{Demand}_t)_e$
COST MODEL
Custom S/W
Ops & Maint
COTS H/W & S/W
Life Cycle Cost
Operational Requirements
BOM
PRODUCTION MODEL
AHWGP
ATBD_i
$(\text{Inst}_1\ \text{Process}_1)_q$
$(\text{Inst}_1\ \text{Process}_2)_q$
$(\text{Inst}_n\ \text{Process}_m)_q$
Sensitivity Analyses
Variations
Results
System Engineering
Design Changes
SLOC Estimate
Price Information
Release Schedule
Release Content
Resource Consumption

EXHIBIT 2-1: The ECS Modeling Context

The User Model describes the interaction of Earth-science researchers and other users with the EOSDIS (the "pull" model). This model is intended to predict what the user needs and loading will be for the EOSDIS and how the demands will vary, over time, as more capabilities come on-line and users gain more experience with the system. The different types of users (see Exhibit 2-2) have been surveyed with somewhat limited utility so far. A new survey, the EOSDIS & General Science User Survey (EGSUS), is being distributed by HAIS, via MOSAIC, to better identify what set(s) of science data products—from a list of products available over time—are of interest to the users and how and to what extent they may interact with the EOSDIS. The users' stated interaction scripts are compiled into scenarios (see Exhibit 2-3). The results of this compilation are then translated into user services (see Exhibit 2-4) that form the basis for the Performance Model's "pull" workload characterization. The temporal details are derived from the relative access frequency distributions of services (as indicated in the scenarios) within epochs of time (see Exhibit 2-5) that closely map to mission milestones. These spreadsheet- based workload distributions are finally translated, manually, into the Performance Model's input parameter tables. As the ECS matures and current Version 0 (V0) user-interaction measurements become available, it should be possible to calibrate the model with real-world information and generate workload predictions with progressively improving confidence.

The ECS Contractor View: Traditional Disciplines	**The USGCRP View† Global Change Research Areas**
Atmosphere (atm)	Climate and Hydrologic Systems (chs)
Cryosphere (cryo)	Biogeochemical Dynamics (biodyn)
Land (land)	Ecological Systems and Dynamics (eco)
Ocean (ocean)	Human Interactions (humint)
	Earth System History (hist)
	Solid Earth Processes (solid)
	Solar Influences (solar)

† U.S. Global Change Research Program

EXHIBIT 2-2: The EOSDIS Users (Researchers)—Two Perspectives

Scen. No.	Scenario Description	Science Discipline	USGCRP Research Area
1	Ph.D. student needs information for dissertation literature review	----	-----
2	Researcher studying lightning associated with flash floods	atm	chs
3	Test ecological theory regarding vegetation competition in grasslands across the central United States	land	eco
4	International researcher (Scotland) developing forest model	land	eco
5	Earth-science researcher wishes to access electronic journal	----	---
6	Regional park land management	land	eco
7	Development of method to integrate data sets of varing resolutions	land	eco
8	Study of biomass burning	land	eco
9	Undergrad. in remote sensing class needs info on EOS instruments and data sets	land	eco
10A	Land-surface hydrologic model	land	chs
10B	Validation of cloud properties with field data	atm	chs
11A	Arctic ice pack response to weather	cryo	chs
11B	Derivation of snow water equivalents	cryo	chs
11C	Radiative fluxes over sea ice	cryo	chs
12	Mid-latitude and tropical interactions—precipitation forcing	atm	chs
13	Earth-science community user; e.g., university prof., radiation budget	atm	chs
14	Development of automated snow mapping procedure (Sequoia 2000 scenario)	land	chs
15	NOAA researcher studying seasonal and diurnal variation in regional lightning distribution	atm	chs
16	Southern ocean large scale circulation	ocean	chs
18	Watershed modeler updating model inputs and providing output to the EOSDIS	land	chs
19	Biogeochemical fluxes at the ocean/atmosphere interface	ocean	bio
20	ISI global water cycle; includes model verification through field studies	land	chs
22A	Thermal alarm system for detection of volcanic eruptions	land	solid
22B	Climatic and tectonic processes in the Andes mountains	land	solid
23A	Stratospheric chemistry and dynamics	atm	chs
23B	Validation of passive microwave algorithm for precipitation retrieval	atm	chs
24	EOS instrument investigator; e.g., MODIS, ocean color	ocean	chs

Scenario distribution according to...
Science discipline: atm-7, land-12, cryo-3, ocean-3, no pref-2
USGCRP area: chs-16, biodyn-1, eco-6, solid-2, hist-0, humint-0, solar-0, no pref-2.

EXHIBIT 2-3: The EOSDIS User Scenarios

User Service Class	Number of Discrete Functions/Services
Search	4
Manipulate	21
Inspect	13
Archive	6
Ingest	1
Produce	1
Other	3
Total	49

EXHIBIT 2-4: The EOSDIS User Services

Epoch	PDR Priority	Date Coverage	Mission Events
A	High	Dec 96 - Jun 97	IR-1, V0 Dataset Migration
B	High	Jul 97 - Dec 97	TRMM (CERES, LIS, VIRS, PR, TMI)
C	High	Jan 98 - Jun 98	EOS AM-1 (ASTER, CERES, MISR, MODIS,
D	Low	Jul 98 - Dec 98	Landsat 7 (ETM+)
E	High	Jan 99 - Jun 99	ADEOS II, CNES or GFO,ACRIMSAT
F-L	n/a	Jul 99 - Dec 02	(Out of PDR Scope)

EXHIBIT 2-5: The EOSDIS Epochs

The Production Model describes the science product generation processes (nominally, Level 0 input transformation to Levels 1/2/3/4 products—the "push" model). This model is intended to predict the steady-state and exceptional processing necessary to deliver trusted science data to the EOSDIS archives when required by the users. The inputs to the model are currently derived from the work being performed by the Ad Hoc Working Group for Production (AHWGP), chaired by Dr. Bruce Barkstrom, and the Algorithm Theoretical Basis Documents (ATBDs) generated by each of the EOS instrument teams (see Exhibit 2-6). The ATBDs describe the scientific rationale for each discrete product. These inputs are translated into sets of science data production process characteristics for each instrument. Each EOSDIS epoch is subdivided by the quarter-year, into which the applicable processes are assigned. This information forms the basis for the Performance Model's "push" workload characterization, over time: data arrival rates and volumes, process/archive physical location(s), process sequencing, inter-product dependencies, quality control, algorithm integration and test; and the computational, transient/archive data storage, and data transport requirements. These spreadsheet-based workload distributions are finally translated, manually, into the Performance Model's input parameter tables. As the ECS matures, it should be possible to calibrate the model with real-world information and generate workload predictions with progressively improving confidence.

Mission	Instruments or Equivalents
TRMM	CERES, LIS, VIRS, PR, TMI
EOS AM-1	ASTER, CERES, MISR, MODIS, MOPPITT
Landsat 7	ETM+
FOO	COLOR
ADEOS II	Sea Winds
ALT RADAR (CNES or GFO)	AMR, DORIS, SSALT
ACRIMSAT	ACRIM
DAO	R. Rood Products
V0	Migrated Datasets

EXHIBIT 2-6: The Current Modeling Scope of EOS Instruments

The Performance Model is intended to provide a definitive basis for evaluating alternative ECS architectures capable of supporting user and production demands (as predicted by the User and Production Models), and to evaluate architectural sensitivities to predictive uncertainties. This model is implemented using the Block Oriented Network Simulator (BONeS) discrete-event simulation modeling tool. A top-level description of the current model under development is illustrated in Exhibit 2-7. As the model scope matures, subsystem models (currently, a mix of static and dynamic models) may be used to supply response characteristics. Doing so, isolates their implementation details and mitigates the system-level model's execution resource demands, which tend to be extensive. The extent to which this may be done is still to-be-determined.

The Performance Model, when complete, should be capable of yielding several important categories of information by:

- Providing resource consumption statistics which could be used by User and Production modeling personnel to assess the impact and improve the performance of their processes;
- Identifying the driving parameters (i.e., the "tall poles" of a histogram) coupled with the capability of evaluating architectural sensitivities to their values in order to help focus analyses at minimizing their uncertainty;
- Assessing expected performance requirements compliance by producing response-time and other statistics that can be compared directly to the requirements;
- Establishing a firm basis for a bill-of-materials (BOM) and a set of operational requirements (i.e., media handling, etc.) necessary to implement, maintain, and operate the ECS for each epoch under consideration; and
- Supporting the assessment of performance vs. cost impacts for new technologies under consideration for incorporation into the ECS.

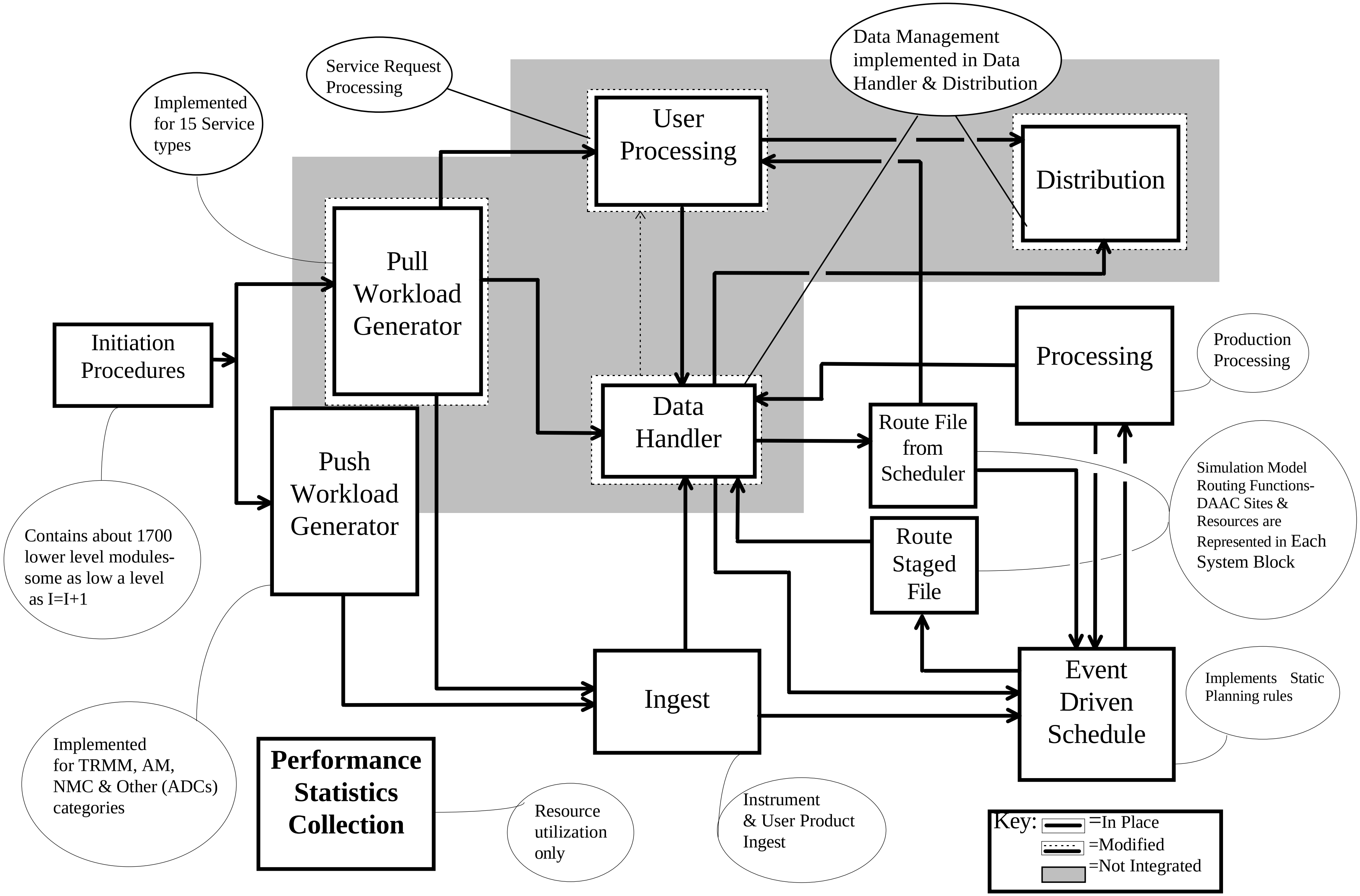


EXHIBIT 2-7: Major Components Of The BONeS Performance

As the ECS matures, it should be possible to calibrate the model with real-world information and generate performance predictions with progressively improving confidence.

The Cost Model is intended to estimate the resources required to develop and operate ECS architecture alternatives (as partially derived from the Performance Model) within schedule constraints. The Cost Model is currently implemented as a collection of three types of stand-alone cost estimation models: Custom Software, Operations and Maintenance, and Commercial-Off-The-Shelf (COTS) Hardware and Software. There is no integrated Lifecycle Cost Model. Each model type, itself, may be composed of several different models or small variations of the same model. The transfer/correlation of information between the models and the aggregation of costs to calculate the overall lifecycle cost are largely manual operations. The IV&V team's visibility into the details of the models is very limited. If the models are as they have been represented to us, as the ECS matures, it should be possible to calibrate them using actual ECS experience and generate future cost predictions with progressively improving confidence.

This IV&V assessment was performed under EOSDIS IV&V Task 5 (Requirements Analysis & Traceability), specifically as part of Subtask 5.3 (EOSDIS User Satisfaction Assessment).

2.5 References

A list of references utilized in these analyses is contained in Appendix E.

2.6 Tools and Data Bases Utilized

A description of the tools and data bases utilized in these analyses is contained in Appendix F.

3. USER MODEL

The User Model (i.e., "pull" model) describes the interaction of Earth-science researchers and other users with the EOSDIS. It is intended to characterize a user profile describing who the users are, what information they need over time, and how they are likely to interact with the EOSDIS. Results of the User Model formulate the basis for determining the inputs to the Performance Model. The findings of an independent assessment of the User Modeling activities, including analysis tasks performed, constraints affecting the analysis, and analysis results, conclusions, and recommendations, are presented in this section.

3.1 Analysis Tasks Performed

In the context of ECS architecture design, it is important to have available an accurate user "pull" model. Information from this User Model is to be used as input to Performance Models in an overall ECS system model. We view the User Modeling activities as representing four major categories:

I	User profile (*who*)	Who are they? What Earth-science discipline and USGCRP research area do they belong to?	Section 3.1.1
II	User needs (*what*)	What data, products, and other information will they need? From which instruments? How much?	Section 3.1.2
III	Temporal distribution of user access (*when*)	When will they need the data? How often will they access the system and for what duration?	Section 3.1.2
IV	Input to Performance Models (*how used*)	How will this information be translated into input to the BONeS Performance Model?	Section 3.1.3

3.1.1 User Profile (*Who*)

Who are the users of the EOSDIS and what are their interests? What has the ESDIS Project done to identify them? What has HAIS done to identify them? What have others, such as Barkstrom [28], done? How does this all fit in with the Global Change Data and Information System (GCDIS) concept of users? The EOSDIS will be a major part of the GCDIS, so it is important that the classes of users selected by HAIS for their modeling work reflect the interests of the GCDIS users. We summarize here the different definitions, or concepts, of EOSDIS users taken from the appropriate sources [2].

The ESDIS definition comprises research users, including U.S. Government-sponsored and other researchers; noncommercial operational- and environmental-monitoring public-sector agency users; applications demonstrations; and others. The last category includes commercial and educational users. The HAIS definition comprises EOS science, general science, and non-science users. Barkstrom considers the Earth science research community; Federal, State, and local government agencies concerned with policy, land use, fisheries, environmental law enforcement, and similar activities; educational institutions; profit-making research organizations and individuals; and interested members of the general public. The Interagency Working Group on Data Management for Global Change (IWGDMGC) considers three categories of GCDIS users:

the general research community, summary seekers of data addressing particular global change problems, and policy makers and planners. This categorization does not specifically point to the EOS-funded users, nor does it attempt to further categorize users along the lines of the ESDIS definition. But then the GCDIS users will be from a broader, more general population that includes many users from the general public as well as global change researchers. There is concern that the general (in the GCDIS sense) or "other" (in the ESDIS sense) users' needs and requirements may not be fully accounted for in the design of the EOSDIS.

We have analyzed four independent data sets, the most important of which are the NASA Headquarters Office of Mission to Planet Earth (MTPE) information data base (a survey containing information about potential EOSDIS users) and the NASA Headquarters EOS directory data base, which contains information on all the EOS-funded investigators—Team Members and Leaders, Instrument PIs and CoIs, and the Interdisciplinary Science PIs and CoIs. The results will be presented according to the USGCRP view of the users' research interests (see Exhibit 2-2).

3.1.2 User Needs (*What and When*)

The central element in HAIS' modeling of user needs is the gathering of information through the creation of a set of user scenarios constructed by interviewing a number of Earth-science researchers. HAIS' goal was to select a set of scenarios that represents the entire ECS user community while having an even distribution from the atmospheric, oceanic, and land disciplines. Not all the goals were met. Nonetheless, the work to date is both substantial and innovative. Their current results are reported in three documents [6,7,8].

We analyzed the scenarios to determine how well they represented both the USGCRP research areas and the traditional Earth-science disciplines listed in Exhibit 2-2. We determined the number of users in each research area or discipline using HAIS-determined science-user demographics for the mid-1999 epoch,[1] which provides the number of users of each scenario.

3.1.2.1 User Scenario Analysis

The science-user requirements for data and services to be requested from the ECS determines the "pull" load on the system, and an accurate assessment of the characteristics of this load is essential as inputs to both the Performance Model and the system design. HAIS has identified 27 user scenarios to represent the science-user requirements categorized by using six styles of system access and four styles of data access—a 6x4 user scenario matrix. HAIS analyzed the scenarios, in discussion with the scientists, and obtained the corresponding ECS functional and service requirements. HAIS produced demographic information (providing the minimum and maximum number of users of the scenarios during several time frames) using the user characterization methodology they developed for this purpose [7].

[1] HAIS made their user scenario matrix spreadsheet, which contains information for four epochs, available to us.

We examined the scenarios to evaluate how well they represent the data and access requirements of the science user community, and we examined the methodology and data used for determining the relative service requirements.

3.1.2.2 Satellite Data Requirements of User Scenarios

The load on the ECS due to user access depends on the user service and the satellite data requirements. HAIS analyzed the user scenarios identifying the steps involved, the functions and services invoked in each step, the satellite data accessed, and the anticipated minimum and maximum number of users for each scenario. We examined these data and mapped the satellite data requirements of all the scenarios to the data expected to be available during the epochs being considered for the PDR to find out if the scenarios represent utilization of most satellite data that will be available during those epochs.

3.1.2.3 Mapping Functions and Services from Scenarios to the ECS Level 3 Requirements

The HAIS user scenario analysis identified 49 functions and services from the user scenarios. These were analyzed and mapped to the nearest applicable ECS Level 3 requirements (see Appendix A). The appendix also includes information about which scenario invokes which functions and services, and the likely number of users of these functions and services; the status of acceptance or implementation of the function as indicated by HAIS [7]; and further IV&V comments.

3.1.2.4 The Requested Requirements Data Base (RRDB)

HAIS created the RRDB in 1993 for collecting, evaluating, and monitoring user requirements. It is available to the public through both Internet and dial-up access. Version 3.0 of the data base is currently in use; the name has been changed to User Requirements Data Base (URDB). The URDB includes functions and services identified in the user scenarios. We accessed the URDB and examined all the requirements that arose from scenario development for the status of their current implementation or acceptance.

3.1.2.5 Distribution of Invoked Services

HAIS conducted a detailed analysis of the scenarios to determine the service types and their relative distribution. They used statistics on the current data system usage at the to project future usage along with the service invocations identified in all the user scenarios. We examined the methodology used by HAIS and found it to be capable of giving representative results. We also analyzed the user scenario data base and examined its consistency and accuracy.

3.1.3 "Pull" Input to the BONeS Performance Model (*how used*)

The studies on user characterization and user-scenario generation and analysis have generated information on the frequency of invocation of different services, requirement of satellite and other data products. The Performance Model requires inputs from the User Model to determine the "pull" loads and their diurnal variation to assess the performance of the system.

The following inputs to the “pull” generator model are required:

- Mean arrival time between transactions (all types) as a function of the time of day
- Fraction of all transactions as a function of the daac
- Fraction of transactions as a function of service (for each daac), and
- Ratio of output volume to input volume as a function of input volume.

Both the methods for generating these inputs and the results need to be verified. Most of this information has not been made available to us. We will assess this part of HAIS’ work when we have been supplied with the necessary information.

3.2 Constraints Affecting the Analysis

Not all the data bases used in our analysis were created for the purpose for which they were used in this study, a validation of user characterization. Nonetheless, the data and information contained in these data bases are independent of both those used by HAIS in their analysis, and they contain additional important information about the user community. Furthermore, they contain valuable information about a broader *potential* user community—those interested in MTPE programs.

The characterization of user demography and demand for products is still in progress at HAIS. An EOSDIS & General Science User Survey (EGSUS) is currently being conducted by HAIS. The results of that survey are likely to effect not only the results of the analyses discussed here, but also the creation of user-model inputs to the Performance Model (BONeS).

Details of the PDR Technical Baseline have only become available to us in January 1995, as part of the community documents catalog on the EDHS WWW server.

There have been time constraints on the IV&V team due to the only recent arrival of much of HAIS’ documentation and data.

3.2.1 Assumptions

Some of the assumptions on which the various user characterizations are based differ. For comparison, they are outlined here.

<u>HAIS study</u>:

- EOS-funded investigators are well defined, and interviews with them provide their requirements.
- The marginal cost of reproduction of data is low enough not to influence user demand.
- Lack of an accurate definition of Level 4 products can lead to underestimating the size of certain user communities.
- *Peterson’s Guide to Graduate Programs in the Physical Sciences and Mathematics, 1994* and Earth science and remote sensing journal article authorship can provide a

representative population sample for the general science (non-EOS-funded) user community.

- The non-science users are only interested in EOS standard products.

IV&V study:

- The set of respondents to the NASA Headquarters survey expressing simultaneous interest in a NASA AO, NRA, EOSDIS, satellite data, and satellite observations is representative the EOSDIS science-user community.
- The USGCRP priority research areas represent EOSDIS data and information areas.

The data we used in this study were not collected with the intention of being used as we have here. We were, therefore, not able to estimate uncertainties for the results obtained from those data that are presented here.

3.3 Results

3.3.1 Discussion of Results

We have examined HAIS' modeling goal of creating user scenarios to make it possible to generate inputs to a Performance Model of the ECS. Starting with HAIS' science-user demography, can they logically create the necessary inputs to the BONeS Performance Model that will meet the user requirements and the system requirements? Do the existing user scenarios contain sufficient information to accurately estimate the "pull" load on the system?

3.3.1.1 User Characterization

The MTPE Data Base. The MTPE office at NASA Headquarters (Code Y) maintains a data base of all contacts who have expressed any interest in any of its programs. This data base contains information on a broad range of potential EOSDIS users: Earth science researchers, other researchers, faculty and students from diverse fields, educators, librarians, policy makers, State and Federal government employees, and so forth. During a recent update of this data base, a survey form was sent to all persons entered in the data base. On this form were checkboxes for the respondent to specify areas of interest. Among the areas available on the form are the categories corresponding to the established science priorities of USGCRP (reproduced here in Exhibit 2-2). There are also checkboxes on the form for those wishing to receive notifications of Announcements of Opportunity (AO) and NASA Research Announcements (NRA) and for those interested in the EOSDIS, satellite data, and satellite observations. There are 9,979 records in the data base.

Unfiltered MTPE Data. Exhibit 3-1 shows the distribution of the number of respondents that indicated an interest in one or more of the USGCRP priority areas. We have used all such records from the survey data base for this exhibit. We take this as being representative of the science preferences of a broad range of the public that have already expressed some interest in the programs of the MTPE.

Filtered MTPE Data. Of the 9,979 respondents, 2,899 (29%) indicated an interest in receiving further information on an AO, 5,052 (51%) an interest in an NRA, 5,286 (53%) an interest in either an AO or an NRA, and 2,665 (27%) an interest in an AO and an NRA. Within this last group, 790 respondents expressed interest in the EOSDIS, satellite data, and satellite observations. We are calling this subgroup of 790 respondents *potential EOSDIS users*—those respondents expressing an interest in an AO and an NRA and EOSDIS and satellite data and satellite observations. Of these potential EOSDIS users, 228 expressed an interest in all seven USGCRP priority areas. Hence, we assume that this last subgroup may have nonspecific science tastes or may be catalog collectors, and have omitted them from further analysis. We have simply selected this reduced group as being representative of those respondents who may be potential EOSDIS science users. Exhibit 3-2 shows the distribution of these users according to science-area preference.

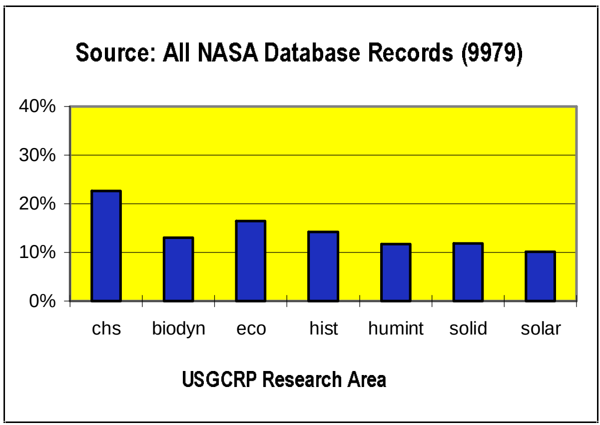


EXHIBIT 3-1: Science Interests Of All Respondents

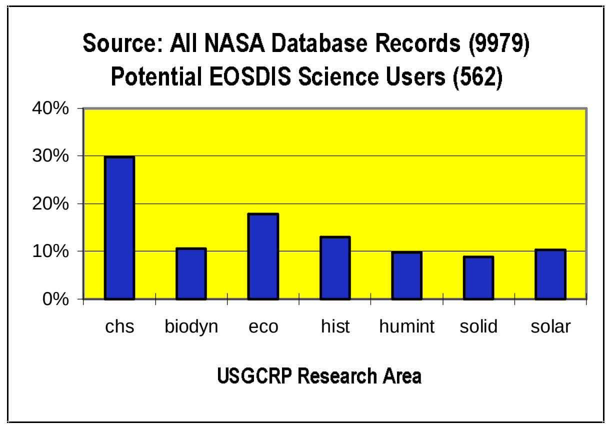


EXHIBIT 3-2: Science Areas Of Interest For The Potential EOSDIS Users

The EOS Directory Data Base: EOS-Funded Investigators. The EOS Directory data base contains, among many other things, information about the EOS-funded investigators. Most of these investigators (555) have also responded to the NASA Headquarters survey, and have stated their USGCRP research area interests and preferences. Those preferences are shown in Exhibit 3-3.

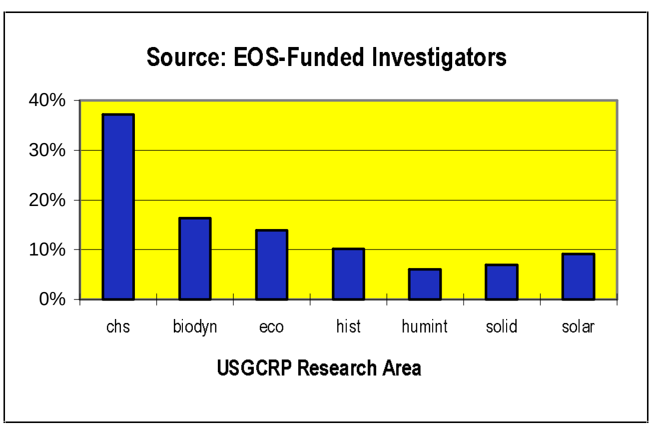


EXHIBIT 3-3: Research Area Interests Of The EOS-Funded Investigators.

Comparison Between User Characterizations by HAIS and this Study. It is clear that several different Earth-science disciplines will have a need for the same data. HAIS has presented a distribution of the expected science users according to Earth-science discipline area. This information was obtained from an analysis of a literature survey of journals and the assignment of the articles to the appropriate Earth-science discipline. We created a similar distribution from the information in the NASA Headquarters survey. (There was no way to extract information from the survey about those interested in the cryosphere, so there is not a one-to-one correspondence between two distributions.)

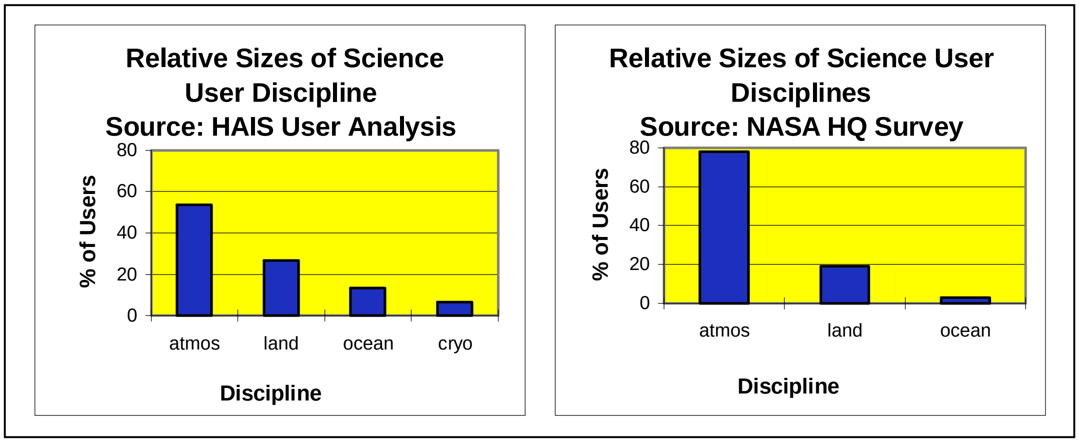


EXHIBIT 3-4: The Distribution Of Science Users According To Discipline.

A comparison of these distributions (see Exhibit 3-4) shows that the relative sizes of the science-user populations (categorized by Earth-science discipline) determined in the HAIS User Modeling effort and the corresponding distribution obtained in this work are similar. The HAIS analysis is based on a broad population obtained from a literature search. Our analysis is based on a broad, but different, set of potential users, and may be taken as an independent confirmation of HAIS' results. It is notable that both analyses show that more than 80 percent of the users are likely to be from the atmospheric and land disciplines, and that those interested in atmospheric questions far outnumber those with land-discipline interests.

3.3.1.2 User Scenario Analysis

We categorized the 27 user scenarios by both traditional science discipline and USGCRP research area. The distributions are shown in Exhibit 3-5; details are in Appendix A.

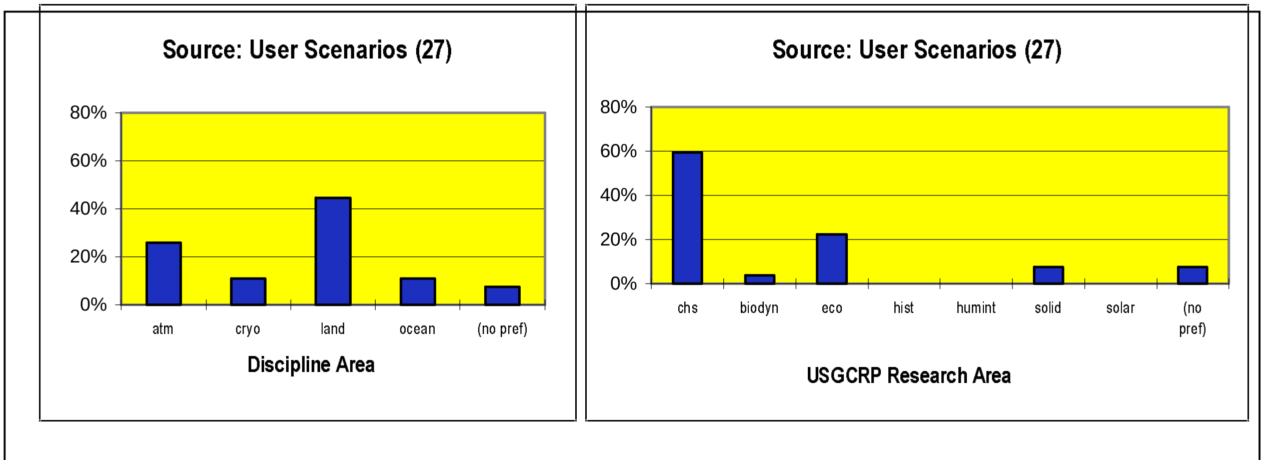


EXHIBIT 3-5: Comparison Of The Science Areas For The User Scenarios.

In the scenario selection process, HAIS' goal was to select a set of scenarios that represents the science user community while having an even distribution from the atmosphere, oceanic and land disciplines. That goal has not been fully achieved. Moreover, the distribution of scenario disciplines matches neither HAIS' nor the NASA Headquarters' expected science-user discipline distribution (Exhibit 3-4).

A similar analysis using the numbers of users of the selected scenarios (for the mid-1999 epoch) is shown in Exhibit 3-6. The numbers of expected users were obtained from HAIS' user scenario spreadsheet; the relative distributions are similar for the other three epochs considered by HAIS. This distribution of users according to disciplines is not uniform (one of HAIS' goals) and like the distribution of scenarios themselves, matches neither HAIS' expected distribution of science users (Exhibit 3-4) nor our analysis of the expected science users (Exhibit 3-2).

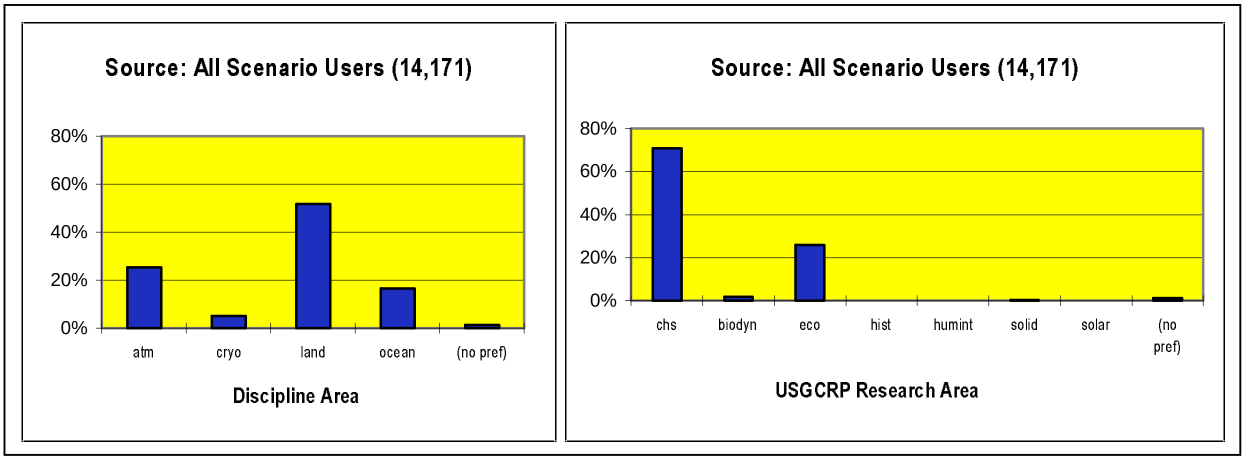


EXHIBIT 3-6: Analysis In Terms Of Number Of Scenario Users

The USGCRP prioritized the research areas in the order given in Exhibit 2-2, which was chosen for the abscissas in the USGCRP-related histograms. The histogram of the EOS-funded investigators' research interests (Exhibit 3-3) essentially follows this priority order.

Instrument Representation in the User Scenarios. The user scenarios were selected to obtain representative model inputs on the access patterns and data requirements during the 1998–2000 time frame. Results of the analysis (Appendix A) indicate that the scenarios have substantial requirements of data from the instruments on AM-1 and Landsat-7 and the AVHRR instrument, and none of the scenarios identified so far require data from ADEOS II, ALTRADAR, and ACRIMSAT, which are expected to be available during 1998–2000 time frame. One scenario requests data from TRMM. It was indicated in the PDR Technical Baseline that the data volume from ALTRADAR and ACRIMSAT is small.

Function and Service Analysis. We mapped the 49 functions or services which were identified from the 27 user scenarios to the ECS Level 3 requirements (see Appendix A). As the functions or services are generally low-level requirements, it is difficult to find a matching functional requirement at Level 3, so they are mapped to the nearest applicable ECS Level 3 requirements. Our analysis shows that 19 of the 49 identified functions or services could not be mapped to the Level 3 requirements. Even though many of the functions in such areas search and subsetting are mapped to the Level 3 requirements, details of the requirements are reported by HAIS to be under differing stages of evaluation.

The URDB Process. Although the URDB process is potentially very useful for collecting and evaluating user requirements, we see that entries in the URDB by persons from states other than Maryland is small (see Exhibit 3-7). This is because a large number of entries in this data base were entered by HAIS from the scenario collection process, accounting for the large number of entries from Maryland. The usefulness of the URDB as a requirement collection device could increase if its existence and utility is well publicized to make it representative of the scientific community throughout the United States. Of the 49 functions collected and analyzed as part of the scenario development process, 34 are currently under differing stages of consideration and 15 have been either closed or rejected (see Exhibit 3-8). There is a need for speeding up the process.

State	Number of Entries
AK	7
CA	4
CO	13
MD	562
MN	2
SD	23
VA	2

EXHIBIT 3-7: URDB Entries By State

Status	No. of Requirements
Being Assessed	9
Closed (comment)	2
Closed (existing req.)	7
Closed (new req.)	0
Closed (non ECS)	2
Design Consideration	12
Reexamine	6
Rejected	4
Screening	7

EXHIBIT 3-8: Status Of Requirements From User Scenarios In The URDB

Distribution of Service Types. Information about the frequency and diurnal variation of invocation of service types is a required input to the Performance Model. For PDR, HAIS has selected 15 service types invoked in the user scenarios and computed their relative distribution by estimating the number of times each service is invoked in a day and multiplying this number for each step of the scenario by the number of users of the scenario and the number of times the scenario is repeated in a year (information available in the user scenario data base). HAIS developed a methodology for determining the expected distribution of service types and access rates represented by the user scenarios. The information available to us indicates that the methodology, which is being refined, is adequate. However during the analysis, we noticed some inconsistencies in the user scenario data base regarding the average number of times services are invoked per year and the user demographics. The matter was discussed with HAIS; they have also found some inaccurate entries and updated the data base.

The analysis identified that some scenario-requested services involving user processing outside normal data product generation, browse, and visualization are being examined by HAIS. These services are important to the science users, and acceptance of these requested services will increase user satisfaction with the system. These are

	Requirement	Status
1.	Identification of user-recommended product processing that is beyond the area of basic product generation in the ECS. (Ability to access, view, manipulate, and store data without the user committing resources) Requirements for ECS to support APIs developed by users	The outcome of the user-supplied processing trade study and policy decisions will effect the final response to the recommendations. Results scheduled for CDR
2.	Requirement for single-request data ordering for past and future data	This is considered to be more an implementation detail than a requirement and the ultimate disposition will await finalization of the ordering capability to be provided by the ECS.
3.	Perform content-based search in the absence of the content in the metadata	User-supplied methods trade study and science software direct access to data server trade study PDR/CDR
4.	Users have identified several visualization utilities that would be advantageous for their research.	Any visualization functionality beyond basic browse will depend on the results of the visualization trade study due at the PDR

3.3.1.3 "Pull" Inputs to the BONeS Simulation

The Performance Model requires inputs from the User Model to determine the "pull" loads and their diurnal variation to assess the performance of the system. Details of the methodology used for the generation of these inputs are not yet available to us. HAIS will create a white paper describing the details of the methods used as well as the results. They are also developing a spreadsheet to facilitate an analysis of the service invocations and data volumes, as well as to perform other operations on this data base. In the absence of detailed information about the methods used for providing inputs to the "pull" generator, we are unable to comment on their adequacy or the quality of their approach. The current version of the data base has been made available to the us (see Appendix F), and some identified discrepancies were rectified in discussions with HAIS. One of the major inputs required for the "pull" generator is the Relative Product Access Frequencies (RPAF), which are expected to be available as a result of the EGSUS.

3.3.2 Identified Problems

The user scenarios do not adequately represent the expected distribution of science users. The scenarios should be representative of the science-user interest so that the user service and data requirements can be correctly estimated and used as inputs to the Performance Model. The scenarios do not adequately represent the prioritized research areas of the USGCRP. Moreover, they do not request data from ADEOS II, ALT RADAR, and ACRIMSAT, which are expected to be available during 1998–2000. Even though it was indicated in the PDR Technical Baseline that the volume of data from ALTRADAR and ACRIMSAT is small, there should be scenarios using data from these satellites as well as ADEOS II in order to identify additional functions or services that may be related to products from those sources.

A large number of functions or services identified from the scenarios (19 of 49) could not be mapped to the ECS Level 3 requirements, and many of these scenario-based functions that we mapped to the closest applicable Level 3 requirements are in different stages of evaluation. (See, for example, URDB records 525, 609, 611, 613, 618, 622, 626, 627, 629, 630, 632, 634, 637, 639, 643, and 644). Considering that the Performance Model inputs regarding the access rates to different services are based on the availability within the ECS of the functions or services identified in the scenarios, decisions about these requirements should be made soon. What will be the impact on the system if some of the requested functions or services are not provided? What will be the corresponding impact on the science?

3.3.3 Potential Issues

Users have identified several visualization utilities that would be advantageous for their research, and the corresponding functionality to be provided in the ECS is currently under study by HAIS. The results of this study are expected by PDR, and the decisions on the visualization utilities should be expedited.

Details of the browse data products generated are not available.

3.4 Conclusions

3.4.1 Technical Integrity

The User Modeling activities analyzed in this report can be broadly grouped into four categories: user characterization, user scenarios, user services, and product access requirements. Based on the analysis conducted, the technical maturity of the User Modeling activity is assessed using the maturity metrics specified in Exhibit 1-1. The assessments are given below.

User Characterization. There is substantial agreement between HAIS' and our analyses of user demographics—*who* are the users. Moreover, HAIS has reasonably estimated the numbers of science users and nonscience users that may be expected to use the system. Nonetheless, more work is being done by HAIS in this area, notably the ongoing EGSUS survey, which will provide more specific information about user demography. For this reason, we feel that this activity has achieved a somewhat limited maturity, and we look forward to the results of the EGSUS.

User scenarios. The purpose of the generation and analysis of the user scenarios is to assess the science user requirements of data and services provided by the ECS. The scenario selection process was based on extensive preliminary work and a number of assumptions. HAIS modelers had detailed discussions with the scientists during the process of scenario selection and analysis. Even though the scenarios are representative of the data access and the system accesses patterns of the EOSDIS science user (the 6X4 Science User Scenario Matrix), they are not representative of either the HAIS-determined science discipline or the USGCRP research areas, as we have shown. The land science discipline scenarios and number of users far outweighs the atmospheric science scenarios and number of users; this is inconsistent with both HAIS' and our estimates of the user demography. Even though the Technical Baseline refers to both U.S. and international science users, the scenarios do not reflect the requirements, if any, of science users outside the United States. All of these shortcomings will affect the rest of the User Modeling results that depend on an accurate assessment of the user needs. This activity has achieved somewhat limited maturity.

User Services. One of the principal objectives of the user scenario collection and analysis processes is to identify the user-required functions or services that depend on the type of access to the ECS, the data access patterns (local, regional, or global), and the research area. HAIS analyzed the user scenarios to gather information on the scenario-based requirements for service types, frequency of access, anticipated users, and so forth. Although the data collected is highly representative of the scenarios so far collected, its usefulness is limited since the scenarios are not fully representative of the expected user demography. Many of the requested functions or services are currently are in differing stages of evaluation for incorporation into the system. The user services part of the User Model has reached somewhat limited maturity.

Product Access Requirements. Accurate assessment of user requirements for data and products from EOSDIS is an important input into the Performance Model. The user scenarios provided the data product requirements for the specific research scenarios. The EGSUS will provide further information about the science user requirements for data access and delivery in the 1998–2000 time frame. Information on the access frequency and the requirements of the data products (Level 0 to Level 4) will provide additional inputs to the Performance Model. The product list for the survey includes data from satellites expected to be launched by the end of 1998 (TRMM, LANDSAT-7, AM1), but do not include data from ALTRADAR and ACRIMSAT, scheduled for launch in 1999. The EGSUS is not expected to provide any viable inputs to the PDR. This activity has achieved limited maturity.

3.4.2 User Satisfaction

The primary users of the work described and analyzed in this document are the HAIS system designers and Performance Modelers. We cannot at this time determine whether the results of HAIS' User Modeling activities will be completely adequate to meet the Performance Modeling needs because much of the User Modeling work is still in progress. Nonetheless, the User Modeling work to date forms a good foundation for extending the work by incorporating our recommendations, and the improved model may provide sufficient and accurate enough information that can be used as input by the Performance Modelers without reservation.

3.4.3 Trends and Projections

The development of the HAIS User Model has progressed well in the areas of user characterization and functional requirements identification. It is now capable of providing limited inputs to the Performance Model. Considering that the User Model is being subjected to our analysis for the first time now, it is not possible to comment on the trends in the development of the model.

3.5 Recommendations

Based on the analysis conducted to date, we offer the following recommendations:

1. Create and analyze as many additional scenarios as is necessary to:
   - Give an adequate representation to all USGCRP research areas commensurate with the demographics of the community interest in these areas,
   - Be representative of the utilization of data from all EOS instruments available during the identified epochs, and
   - Truly reflect the international character of the science users by including scenario(s) from scientists or institution outside the United States.
2. Update the User Model based on the results of the planned survey (EGSUS), additional information available as a result of the recommendations of this report. In particular, the EGSUS should provide both the respondent's traditional science discipline *and* USGCRP research area. These items should be required fields in the survey.
3. Include functional or service requirements and "pull" load on the ECS that may arise from the access by the International science-user community—particularly International Partners (IPs).
4. Estimate the (increase in the) "pull" load on the ECS the due to non-EOS and non-satellite data after the expected increase in the scientific activity that will result from the general availability EOS data.
5. Estimate the (increase in the) "pull" load due to the access of satellite data from the International Partners (IPs), through ECS by the U.S. science users, and
6. Assign priorities to and binding of the user-generated requirements (from the user scenarios) to the available resources for each release.

3.5.1 Areas Requiring Further Analysis

Areas requiring further analysis are listed here.

1. Continue the analysis of the user characterization part of HAIS' User Model to further identify *who* the user is in terms of both USGCRP research areas and traditional Earth science disciplines based on the results we expect from the EGSUS.
2. Verify the changes to the User Model made by HAIS based on the recommendations in this report.
3. Verify HAIS' methodology used to derive inputs to the Performance Model. Results from the EGSUS are also expected to be important in this area. We recommend that the methodology

used and the results obtained in the generation of inputs to the "pull" generator be subjected to verification as soon as possible, preferably before CDR.

4. Map the user scenario-generated requirements to the ECS Level 4 requirements.
5. Examine the adequacy of the User Model inputs (as they become available) to the BONeS Performance Model based on feedback from the results of Performance Modeling.

3.5.2 Solutions to Important Problems

1. The user characterization needs to be frozen (for the epochs under consideration) well before CDR, based on the EGSUS-derived updates to the model.
2. Complete the trade studies on user processing, visualization and browse products and take policy decisions where necessary early, so as to provide inputs into the model before the CDR.
3. Expedite evaluation of all functional or service requirements identified from the scenarios, identify the functions or services that cannot be implemented in the ECS, and evaluate the impact of omitting such functions or services.
4. The URDB process is a powerful tool for obtaining new user requirements. However, it is currently being used by a small number of potential investigators and is not representative of the broad user community in the United States. This limitation needs to be corrected by advertising the availability of this facility. Additionally, the URDB process needs to be expedited in order to be effective in providing information on new user requirements for the system modeling and its design.

3.5.3 Risk Management

The information that will be derived from the EGSUS needs to be incorporated into the User Model as soon as possible in order to provide accurate estimates of the "pull" load before CDR.

4. PRODUCTION MODEL

The Production Model (i.e., "push" model) describes the science product generation processes (nominally, Level 0 input transformation to Levels 1/2/3/4 products). This model is intended to predict the steady-state and exceptional processing necessary to deliver trusted science data to the EOSDIS archives when required. The findings of an independent assessment of the Production Modeling activities, including analysis tasks performed, constraints affecting the analysis, tools and data bases utilized, and analysis results, conclusions, and recommendations, are presented in this section.

4.1 ANALYSIS TASKS PERFORMED

Our analysis focused on four aspects of the Production Model: 1) the overall context within which it is being developed, 2) the completeness of the work being performed by the AHWGP, 3) the translation of AHWGP inputs into the HAIS spreadsheets, and 4) the translation of the model data into BONeS Performance Model input parameter tables. Our analyses did not address the scientific correctness of the AHWGP work, which is rightly the responsibility of the various instrument teams, but only the maturity of the process definitions at the time they formed the basis for HAIS PDR estimates. Our analyses were severely limited by the lack of timely information on which to base them.

In furtherance of the background discussion in Section 2.4, at the recommendation of the EOS Payload Advisory Panel and the EOSDIS Advisory Panel, after the EOSDIS System Requirement Review, ESDIS redirected the ECS design paradigm as shown in the following:

From	To
Central Operational Control	Operational Enabling
Data Order and Transmission	Data Publishing and Access
ECS Elements	Service and Data Provider Function.

In order for the ECS development contractor to have access to correct information on the system requirements from the EOS instrument scientists and to ensure a better communications, ESDIS created the AHWGP. The objectives of the working group are the following:

- Provide an interface between Science Community, ESDIS Project and the ECS development contractor to ensure that there is a sound basis for describing: data products, operational scenarios, EOSDIS archival size, computational load, networking properties, and the external interfaces necessary for acquiring non-EOS Data.

- Provide a mechanism for building the infrastructure needed to smooth the exchange of documentation and other descriptions from the science community to the ESDIS Project, the DAACs, and the ECS development contractor.

At the ECS System Design Review (SDR) it became apparent that the ECS development contractor had generated static ECS models which did not adequately represent the dynamics of the ECS. Consequently, ESDIS directed the ECS development contractor to construct a Performance Model (i.e., BONeS) of the ECS using the new paradigm.

At the request of the AHWGP, the EOS AM-1 Instrument Teams provided the ECS modelers with information on input and output file sizes, required processing capability for each process, and product phasing scenarios for each of the standard products that will be generated at various epochs of the system lifecycle.

Our current analysis of the Production Model examines:

- The impact of the mission redefinition on the Production Model;
  Instrument and product availability for Production Model inputs at various epochs (see Exhibit 4-2);
- Data dependencies for product generation; and
- Verification of the translation of the information provided by the AHWGP and from the Algorithm Theoretical Basis Documents (ATBDs) to the parameters needed by BONeS.

4.1.1 Impact of EOS Mission Redefinition on the Production Model

Subsequent to the SDR, several significant changes in the objectives of the project occurred. Consequent to these changes, the Production Model was impacted the following ways:

A 20 percent duty cycle for ASTER on-demand processing, and transfer of all ASTER DAR processing to Japan eliminated an entire function from ECS and decreased the “push” load on the system;

Elimination of CER 10 as a standard product deceased the production load;

Inclusion of TRMM and Landsat 7 products into ECS product responsibility imparted new demands, increasing the ECS functionality requirements; and

The requirement of a 20 percent per year growth in processing and storage requirements with profiles for the phasing of data storage and processing capacities introduced a definitive evolvability requirement for the system.

In August, 1994, Dr. Steve Wharton, EOSDIS Project Scientist and Mr. M. Myers, RDC, created a draft Science Operations Concept and sought community consensus on how to maximize the science benefit for a given data product implementation budget based on the idea of Phased Algorithm Implementation with a baseline data capacity that defines a resource envelope into which the science data products are prioritized. He presented a “Hypothetical EOSDIS Capacity Allocation Scenario of TRMM and AM-1” to the Investigator Working Group (IWG) workshop on October 20, 1994 in which he presented a processing load and data volume capacity allocation baseline for TRMM and AM-1 for *at-launch* and *launch plus 2 years* periods. In order to create this baseline, Dr. Wharton estimated a resource capacity envelope shown in Exhibit 4-1.

Period	Processing Load (MFLOPS)	Data Volume (GB/day)
Launch	5,093.30	723.32
Launch Plus 2 Years	29,960.61	1,701.94

EXHIBIT 4-1: Estimated Capacity Envelope

They also created a distribution for the allocation of processing capacity and data volume for various instruments for TRMM and AM-1 missions. Our analysis compared the allocated processing and data volume capacities with the load requirements generated by AHWGP and those used by HAIS in the PDR Technical Baseline (as of December 21, 1994).

4.1.2 Instrument and Product Availability Effects on Production Model Loading

We have examined the HAIS Production Model inputs from the perspectives of availability of instruments at a particular epoch and the requirement of generation of standard products for that epoch. The Mission Baseline considered by HAIS as given in the PDR Technical Baseline is shown in Exhibit 2-6.

Consistent with the above baseline, HAIS has defined several epochs in the ECS system lifecycle. These epochs are related to major events or releases. Exhibit 4-2 shows the epochs and the events related those epochs. A priority number has been attached to these epochs to indicate the urgency to create the required ECS capability during that time frame. In the Exhibit, the ECS responsibilities are shown in ***bold italics*** and the underlined activities are items that HAIS will emphasize at PDR.

Epoch	Priority	Date	Event
A	1	***Dec 96 - Jun 97***	***Release IR-1*** (Dec '96), ***V0 migration complete*** (Mar '97)
			<u>PDR focus: IR-1 w/subset of V0 data + launch support</u>
B	1	***Jul 97 - Dec 97***	TRMM (CERES, LIS, VIRS, PR, TMI) (Aug, '97)
			CERES, LIS: Product Generation, Archive, Distribute
			VIRS, PR, TMI: Archive and Distribute
			TRMM Release (Dec '97)
			<u>PDR focus: IR-1 w/TRMM</u>
C	1	***Jan 98 - Jun 98***	EOS AM-1 (ASTER, CERES, MISR, MODIS, MOPITT)
			ALL INSTR: Prod. gen., Archive, Distribute (June '98)
			<u>PDR focus: Release AM-1 w/ TRMM and EOS AM-1</u>
D	3	***Jul 98 - Dec 98***	Landsat 7 (ETM+) (Dec '98)
			ETM+: Archive and Distribute
E	1	***Jan 99 - Jun 99***	ADEOS II (SeaWinds) (Feb '99)
			SeaWinds: Product Generation, Archive, Distribute
			CNES or GFO-ALTRADAR (AMR,DORIS,SSALT)
			(Mar '99)
			ALL INSTR: Product Generation, Archive, Distribute

Epoch	Priority	Date	Event
			ACRIMSAT (ACRIM) (June '99)
			ACRIM: Product Generation, Archive, Distribute
			PDR focus: multi-mission support
F	3	Jul 99 - Dec 99	Not Defined
G	3	***Jan 00 - Jun 00***	CNES or RSA mission (SAGE III) (Jan '00)
			SAGE III: Product Generation, Archive, Distribute
			Space Station (SAGE III) (June '00)
			SAGE III: Product Generation, Archive, Distribute
H	2	***Jul 00 - Dec 01***	EOS PM-1 (AIRS, AMSU, CERES, MIMR,MODIS, MHS)
			(Dec '00)
			ALL INSTR: Product Generation, Archive, Distribute
I	4	Jan 01 - Jun 01	Not Defined
J	2	Jul 01 - Dec 01	Not Defined
K	4	Jan 02 - Jun 02	Not Defined
L	3	***Jul. 02- Dec 02***	CHEM (HIRDLS, MLS, CII, TES) (Dec '02)
			HIRDLS, MLS, TES: Product Generation, Archive,
			Distribute
?	?	Jul 03	ALT LASER (GLAS)

EXHIBIT 4-2: Details Of EOSDIS Epochs

The Science Operations Concept by Wharton and Myers described a way of baselining, supporting, configuring and delivering science data products through a process which allows the scientists to prioritize the system and allocate a baselined capability to products most in demand, most ready and best performing in the system. In support of this operations concept, the AHWGP provided Product Availability Scenarios for AM-1 instruments. These scenarios are shown in Appendix B. They also provided phasing of processing and volume loads for each product at various epochs (in quarterly time intervals from launch to year 2001) to assist in the development of the ECS Production Model. The HAIS PDR Technical Baseline Attachment C in EDHS Community Access Internet Server contains the following documents (Excel spreadsheets) describing these phasing scenarios: Processing Timelines, Volume Timelines, File Descriptions, and Processing Descriptions (see Appendix F for specific versions).

We have investigated the adequacy of the Production Model inputs in terms of availability of various instrument data at different epochs and analyzed consistency of data in the Excel spreadsheets with those in AHWGP Product Availability Scenarios (Appendix B) and the ATBD requirements. Since the data definition for MOPITT is at a more mature state than most other instruments, we have chosen to focus on MOPITT data as a representative precursor to subsequent analyses. The analysis for MOPITT is presented in Appendix B.

4.1.3 Data Dependencies

There are two types of data dependencies associated with product generation within EOSDIS:

- Data generated in one DAAC required for production purposes in another DAAC.
- Data that must come from sources outside of the ECS.

Several DAAC to DAAC data transfer requirements are shown in EXHIBIT 4-3.

Product Processing	**Processing Site (DAAC)**	**Required Data**	**Dependent Data Originator (DAAC)**	**Data Transfer Requirement**
MODIS Land Products	EDC	MODIS	GSFC	GSFC to EDC
MODIS Land Products	EDC	MISR	LaRC	LaRC to EDC
MODIS Atmospheric Products	GSFC	MISR	LaRC	LaRC to GSFC
MODIS Atmospheric Products	GSFC	MODIS	EDC	EDC to GSFC
CERES Products	LaRC	MODIS	GSFC	GSFC to LaRC
MISR Products	LaRC	MODIS	GSFC	GSFC to LaRC
MOPITT Products	LaRC	MODIS	GSFC	GSFC to LaRC

EXHIBIT 4-3: DAAC To DAAC Data Dependency

An example of external data sets requirements is shown in Exhibit 4-4 for some MOPITT products.

Product ID/Level (DAAC)	Product Name	Process ID	Dependent Parameter Name (ID) from external sources	When Needed (epoch)	Dependent Parameter Level	Type of data	Source	IV&V Comments
MOP042 LaRC	CO Total column burden	This process has not been addressed by the instrument team	Radiosonde Temperature Profile Humidity Profile	VPGI 3/98 VFGI (4/98-1/99) AFGC (2/99-4/02) (IV&V note: Although specified by AHWGP, this product has not been incorporated in the model)	3	P	NOAA NMC/NOAA	MOP04 product has not been included in modeling
MOP032 LaRC	CO Profile	Level 3 Processing (MOPL3) Level 3 QA and Error Analysis (MOPL3Qi-E) Level 3 QA and Error Analysis (MOPL3Qi-F)	Radiosonde Temperature Profile Humidity Profile (IV&V note: as per AHWGP, for MOP1 and MOP2 processing, the following data sets will be required: ANC_EDC_DEM ANC_NMC_PROF ANC_NMC_SURF MOD06_L2_G MOD30_L2_G)	VPGI 3/98 VFGI (4/98-1/99) AFGC (2/99-4/02)	3	P	NOAA NMC/NOAA	

*

N Standard product not available	V Undergoing validation, users beware	A Available for general use
P Partial parameter generation	F Full parameter generation	
R Regional coverage, such as for algorithm proving	I Intermediate, moderately regular coverage	G Global coverage
S Sporadic: only a few, irregular times in a month	I Intermittent: regular, Moderately frequent sampling	C Continuous: Large fraction of possible samples taken
X Designation not applicable for that field		

EXHIBIT 4-4: External Data Dependencies For MOPITT

We have evaluated data dependencies within EOSDIS from the perspective of the requirements needed to form a reliable Production Model workload. A discussion of maturity of inclusion of data dependency in the model is discussed in Section 4.4.

4.1.4 Verification of Spreadsheet Translation to the BONeS Parameter Tables

In addition to the Excel spreadsheets containing processing descriptions and timelines, input/output file descriptions and data volume timelines, each Instrument Team provided the ECS modeling team with system process flow diagrams identifying the various inputs, processes, and outputs for each product processing step. The ASTER process flow diagram is shown in Exhibit 4-5. We have integrated all ASTER product processing in this diagram to show input-process-output involved in the production of all ASTER products. We have investigated the consistency of data from the various sources: Excel spreadsheets, process flow diagrams, and BONeS model input data.

4.2 Constraints Affecting the Analysis

The primary constraints affecting our analysis of the Production Model were: 1) very late access to the HAIS PDR Technical Baseline, and 2) instability of the process definitions.

While the IV&V team had access to AHWGP data on the AHWGP Internet Gopher server, it was not until January 17, 1995 that community access was granted to the HAIS PDR Technical Baseline. Since a substantial amount of information received from the AHWGP was not put on the AHWGP server, it was impractical to proceed with detailed IV&V analyses prior to that date. In addition, HAIS and the instrument scientist interactions produced many modifications to the input data. This resulted in severe configuration issues (i.e., the identification of which sets of data upon which to base the analyses). To achieve some degree of success for this TAR, we have applied our methodology primarily to the Production Model input data associated with MOPITT and (to a lesser degree) ASTER, which are the most mature examples for which we received data at that time.

4.3 Results

4.3.1 Discussion of Results

Some of the results and problems discussed in this section are provisional and preliminary. Since the version of the Production Model analyzed here is not an integrated model and no results have yet been obtained, this TAR does not contain analysis of the model results. The main results of this analysis are:

- The provisional ESDIS capacity allocation for various missions/instruments are not always in agreement with those estimated by the AHWGP and/or HAIS;
- The Production Model does not incorporate all capabilities required by ECS at a particular epoch; and
- Some discrepancies exist among data provided by AHWGP and those used by HAIS in the Production Model.

We believe that HAIS modelers have an understanding of the modeling process and are proceeding in the right direction in this effort.

4.3.2 Identified Problems

Exhibits 4-6 and 4-7 examine the consistency between the processing loads and data volume allocations by Dr. Wharton, those stated by the AHWGP, and those used by HAIS for production modeling.

Platform	Launch Date	Processing Load (MFLOPS)						
		Instrument	At Launch			Launch Plus 2 Years		
			AHWGP Requirement	PDR Tech Baseline (HAIS)	ESDIS Allocation	AHWGP Requirement	PDR Tech Baseline (HAIS)	ESDIS Allocation
		CERES	905.93	905.93	1,139.51	2,779.29	2,779.29	3,014.6
TRMM	Aug. 1997	LIS	0.42	0.42	0.42	0.42	0.42	0.42
		Total	906.35	906.35	1,139.93	2,779.71	2,779.29	3,015.02
		ASTER	99.16	99.16	17.4	99.16	99.16	17.4
		CERES	2,937.01	2,031.08	1,139.51	2,779.29	2,031.08	3,014.6
AM-1	Jun. 1998	MISR	244.99	3,459.53	1,127.82	3,459.53	3,459.53	4,057.0
		MOPITT	3.45	17.66	17.66	17.66	17.66	17.7
		MODIS	3,553.79	3,569.31	2,730.00	3,569.31	3,569.31	8,000.00
		Total	6,838.4	9,176.74	5,032.39	9,924.95	9,176.74	15,106.7

Exhibit 4-6: Processing Load

Platform	**Launch Date**	**Data Volume (GB/day)**						
		Instrument	At Launch			Launch Plus 2 Years		
			AHWGP Requirement	PDR Tech Baseline (HAIS)	ESDIS Allocation	AHWGP Requirement	PDR Tech Baseline (HAIS)	ESDIS Allocation
		CERES	5.29	11.35	26.73	22.62	11.35	27.8
TRMM	Aug. 1977	LIS	1.33	1.33	1.41	1.24	1.33	1.4
		Total	6.62	12.68	28.14	23.86	12.68	29.2
		ASTER	99.16	118.81	59.24	99.16	118.81	59.2
		CERES	23.25	11.9	26.73	22.62	12.22	27.8
AM-1	Jun. 1998	MISR	65.96	136.34	84.05	136.39	136.34	136.4
		MOPITT	0.12	0.19	0.19	0.19	0.19	0.2
		MODIS	552.86	537.23	541.00	633.73	938.25	1,082.0
		Total	741.35	804.47	711.21	892.09	1,205.81	1,305.6

Exhibit 4-7: Volume Load

Processing load estimates for TRMM by HAIS and AHWGP are the same for both at-launch and launch plus 2 years periods. ESDIS resource allocation is adequate to cover the requirements of TRMM processing. However, the data volume estimates are widely different for the three sources.

For the AM-1 at-launch time frame, there are significant differences among the three processing load estimates. The ESDIS allocation varies significantly from either the AHWGP or HAIS estimates. Considering that HAIS' source for this data is the AHWGP, it is not clear as to why the HAIS estimate is so much larger than that of the AHWGP. On the other hand, the volume estimates are reasonably close.

For the AM-1 launch plus 2 years period, while HAIS processing load estimate is lower than that of AHWGP, the reverse is true for data volume estimates. ESDIS processing power and data volume allocation are significantly higher than the other two. The EOSDIS capacity envelope estimates include only a subset of ECS requirements. For example, it does not include processing and volume requirements for Landsat-7 ETM+, TRMM VIRS, TMI and PR , ADEOS II SeaWinds, COLOR, ALTRADAR, ACRIMSAT, SAGE III instruments, migrated V0 data, and DAO data production.

With regard to epochal considerations, the Production Model received by the IV&V Team as of January 26, 1995 does not fully account for all events that are expected to occur during the four epochs of PDR emphasis, namely: A, B, C and E. V0 migration is planned to be completed by March 1, 1997 but only a subset of V0 data is being considered at PDR. In addition to ADEOS

II, ACRIM data products, ECS will create and distribute products from ALTRADAR instruments: AMR, DORIS and SSALT during Epoch E. Product processing, archive and distribution loads for these products have not been incorporated into the HAIS Production Model.

The results of an analysis of consistency of AHWGP Product Availability Scenarios for MOPITT with data found in the HAIS PDR Technical Baseline tables are shown in Appendix B. We found that the timelines in the PDR Technical Baseline did not always agree with those in the AHWGP product availability scenarios. We also found that the processing power estimates in the Technical Baseline were at variance with those in ATBD. Additionally, MOP04 product processing has not been addressed.

It was apparent in the AHWGP workshop during December 13-14, 1994, that several policy questions must be answered before adequate data dependency inputs can be made in the model. Key among these, involves the segmentation of data to be transferred from one DAAC to another. Although MOPITT was chosen as the methodology example for our verification of HAIS translation of AHWGP "push" data into the Production Model, the version of BONeS that we received contained only ASTER "push" tables. Consequently, we conducted a preliminary analysis of the contents of the BONeS file (F_Desc.txt) containing descriptions of "push" input data for ASTER. We have re-drawn the Instrument Team's ASTER product processes flow diagrams (EXHIBIT 4-5) to show all ASTER product input-process-output in one diagram. A consistency check demonstrates that several discrepancies exist among data used in the ASTER process flow diagrams, file descriptions in the HAIS PDR Technical Baseline, and the BONeS F_Desc.txt file (Appendix B). Examples of these are summarized in EXHIBIT 4-8:

File descriptions missing in PDR Tech Baseline but found in BONeS and Process Flow diagrams	BONeS Files not found in Process Flow Diagrams	Process Flow Files not found in BONeS
AST_14	AST_MODTN3	AST_DEM_GRD_TMP
AST_ANC_MIS05	AST_MODTN4	AST_DEM_PIX_TMP
AST_NMC	AST_DEM_GRD_TMP1	
AST_ANC_TOMS	AST_DEM_PIX_TMP1	
AST_ANC_MIS12	AST_DEM_GRD_TMP2	
AST_ANC_MOD30	AST_DEM_PIX_TMP2	
AST_ANC_MOD30B		
AST_ANC_MOD10		
ANC_ECOSYS_DB		

EXHIBIT 4-8: Discrepancies In Data Files In Various Sources

4.3.3 Potential Issues

The phasing of various product generation sequences require a better understanding of data dependencies for efficient product generation. A small mismatch could cause serious production problems. Also, the definition of the "size and shape" of processing granules for different instruments and the associated problems of data segmentation and distribution responsibilities still remain to be resolved.

In their presentation on Dec. 5, 1994, HAIS stated that “the AM-1 archive will not include Level 0 data products”. However, it is our understanding that, in addition to EDOS, each DAAC will archive Level 0 product data for which it is responsible. It is not yet clear how much of Level 0 archival and distribution load will be an ECS responsibility.

During the December 13-14, 1994 AHWGP workshop, HAIS correctly reported that AHWGP data has not provided the ECS modelers adequate information on:

- Algorithm Integration and Test (AI&T) workload;
- Reprocessing Requirements;
- Ancillary product specifications (external sources not identified, availability/sizing information not validated);
- Error budget for processing/file sizes;
- Data distribution requirements other than Q/A;
- Data dependencies (conflicting assumptions by product developers and users); and.
- AHWGP estimation of the theoretical MFLOPS/PGE execution did not include system overhead and CPU utilization.

The above information must be integrated into the Production Model to create a realistic representation ot the ECS system.

4.4 Conclusions

4.4.1 Technical Integrity

Our major conclusions concerning the technical integrity of the Production Model, as of the time it was intended to support PDR, are:

- AHWGP inputs are still incomplete and unstable;
- Significant issues must still be resolved before a reliable Production Model can be implemented; and
- The translation of Production Model inputs to BONeS is incomplete.

4.4.2 User Satisfaction

The Production Model will have a wide range of users: the designers of the ECS system; the ESDIS project personnel; the Instrument Team members; ECS system Integration and Test organizations; and EOSDIS IV&V personnel (to support system certification), etc.. Since a very immature production model exists at the present time, user satisfaction can only be rated as marginal, at best.

4.4.3 Trends and Projections

The purpose of this section is to highlight measurable differences observed between the results of the current analysis and previous ones; and to project the implications of those differences into the future (i.e., whether they appear to be diverging from, or converging toward a reliable Production

Model. Since this is the first analysis performed by IV&V (i.e., the first datum), it is not possible, at this time, to document a trends analysis or project trends into the future.

4.5 Recommendations

4.5.1 Areas Requiring Further Analysis

The breadth of IV&V Production Model analysis needs to be expanded to include all missions/instruments that play a significant role in the production loading of the ECS. Now that analysis input data is becoming more readily available, this effort should prove to be more illuminating in terms of maturity level quantification. Future IV&V analyses (prioritized by workload impact) can be expected to yield maturity metrics by mission/instrument/process, together with accompanying engineering analysis rationales.

4.5.2 Solutions to Important Problems

We recommend the following:

- Include missing/incomplete 'push' loads. These include unaccounted for product processing, algorithm integration and test workloads, as well as transient and archival sizing requirements;
- Resolve discrepancies among various data sets and correct data translation errors; and
- Create a stable, configuration-controlled, approved ESDIS Production Model baseline as soon as possible.

4.5.3 Risk Management

An important risk factor in the development of an ECS that meets performance goals, is the lack of a full understanding of the "push" workload. Considering the complexity of the ECS, the design must take full advantage of any modeling activities. We recommend that the model implementation process be accelerated, such that model results can be used to reliably identify and rectify problems in the design, well prior to CDR.

5. PERFORMANCE MODEL

The Performance Model is intended to provide a definitive basis for evaluating alternative ECS architectures capable of supporting user and production demands (as predicted by the User and Production Models), and evaluating architectural sensitivities to predictive uncertainties. This model is implemented using the Block Oriented Network Simulator (BONeS) discrete-event simulation modeling tool. The findings of an independent assessment of the Performance Modeling activities, including analysis tasks performed, constraints affecting the analysis, tools and data bases utilized, and analysis results, conclusions, and recommendations, are presented in this section.

5.1 Analysis Tasks Performed

5.1.1 Evaluation Areas

The purpose of this analysis was to evaluate the HAIS BONeS performance model of ECS. Specifically, the evaluations performed are as follows:
Evaluate representation of the system design;
Evaluate the usage of the input parameter values;
Evaluate statistics collection/performance metrics quantification; and
Evaluate overall model structure.

5.1.1.1 System Design Representation Evaluation

The first evaluation area involved assessing the BONeS model representation of the ECS system design, specifically the: SDPS subsystems; the “push” and “pull” workloads; and the distributed architecture resources (e.g., processors, internal I/O channels, storage devices, local networks, and wide area networks).

Our analysis approach was to examine the BONeS model modules for completeness and correctness with respect to the system definition, workload requirements, and preliminary design details. The model should have the SDPS subsystems and a set of workload flows mapped onto a distributed system of computer and communication resources that represent the SDPS and CSMS hardware design. The subsystems represent the ECS application functions at the DAACs, whereas the “push” and “pull” workload execution flows represent the transactions for the instrument data files and the users service requests, respectively. The BONeS modules were evaluated for their representation of the following areas:

- “push” and “pull” workloads;
- Processes;
- Files/Data Sets;
- Subsystems; and
- System resources.

The “push” and “pull” workloads define the transaction sources (e.g., by instruments for “push” workloads and by user services for “pull” workloads). Each workload should be defined in terms of frequency of arrival by epoch, and execution flow through a set of processes at (possibly) different locations in the distributed architecture.

Processes are defined by the processing site and resource, the input and output files, and the resource service demands (e.g., number of floating point operations). Files define the “push” inputs by instrument and the “pull” products by service. Their characteristics should include size, archive site, storage retention (i.e., archive, permanent, interim, or temporary), and other parameters such as temporal and spatial coverage. Subsystems define the functions of ECS. Their characteristics should include services provided, interfaces, workload mapping, system resources mapping, and process mapping. System resources define the hardware servers in the design. Their characteristics include resource type (e.g., processor, disk, I/O channel, network link, robot, juke box, memory, and tape), location, and capacity parameters (e.g., MFLOPS, seek and rotation times, storage size, and transfer rate).

5.1.1.2 Input Parameter Evaluation

In the second evaluation area, the usage of the model input parameters was evaluated. The parameters include the following categories:

- “push” and “pull” transaction arrival rates;
- Service demands for processing, storage I/O, and network transfer; and
- System capacities, overheads, and configuration parameters.

Our analysis approach was to examine the model input files and the parameter values defined within the BONeS modules for appropriate usage in the modules. Parameters were evaluated for “push” and “pull” workloads, processes, files, topology, and system resources. The evaluation of the actual values and their derivations are discussed in Sections 3 and 4 of this document.

The “push” transaction arrival rates and data volumes were derived by HAIS from data provided by the science instrument teams. The “pull” transaction arrival rates were derived from the user scenarios developed by HAIS. The information for the user scenarios is documented in numerous HAIS working documents and Excel spreadsheets. These sources and the “pull” workload derivation are discussed in Section 3.

Service demands are the units of service consumed by “push” and “pull” transactions at the different computer and communications resources where they are processed or transferred. These estimates of floating point operations or instructions, file sizes, and message sizes were made by the instrument teams for the “push” workloads and by HAIS for the “pull” workloads.

System capacity, overhead, and configuration parameters are derived from the HAIS design of the distributed architecture. The capacities define the execution rates (e.g., MFLOPS) and transfer rates of the system resources. Overhead factors, or parameters, are performance penalties that represent service demands that are not modeled for a particular resource (e.g., protocol control packets or operating system software). Configuration parameters include details of the design such as numbers of devices and data segmenting lengths for storage and network I/O operations.

5.1.1.3 Output Statistics Evaluation

The third evaluation area involved assessing the performance statistics collection defined in the model. These statistics provide the data for computing system performance metrics such as resource utilization, user response times, and component delays and queue lengths.

The analysis approach was to examine the statistical probe modules and the simulation configuration file in the model to determine what performance metrics were quantified by the model. When output results become available, we will evaluate the statistical soundness of the data analysis methods by determining the collection period for the statistics and the statistical functions performed on the data. We will evaluate the model’s approaches for avoiding system start-up transient affects in the data; for determining steady state conditions; and the simulation execution time which should be long enough period such that the results are statistically significant.

5.1.1.4 Model Structure Evaluation

In the fourth evaluation area we analyzed the overall model structure for the purpose of determining how easily it would be to modify the model to address deficiencies or to add additional levels of detail or functionality.

The previous three evaluations provided the basis for the fourth evaluation. The analysis approach was to examine the overall model structure to determine how subsystems are mapped onto the distributed architecture and system design; how computer/communication resources are represented; how the workloads are mapped onto the resources; and how the input parameters are provided to the model and how they are used within the model framework.

A well-structured model should be able to accommodate modifications in requirements and design without those modifications requiring significant changes throughout the model. The model should also be structured such that statistics collection can be added or modified. If the model is sufficiently modularized, then changes can be isolated, and additions can be made without perturbing other parts of the model significantly. However, this modularization depends not only on the model design but on the capabilities of the modeling tool as well. In some cases, it is not completely possible to prevent dependencies of model components and parameters across multiple parts of the model.

5.1.2 Evaluation Metrics

The evaluation metrics used for the Performance Model assessment are categorized as follows:

- Engineering Quality
- Testability
- Traceability

- Flexibility

5.1.2.1 Engineering Quality Metric

The Engineering Quality metric is defined in terms of the completeness and correctness of the model in representing the ECS subsystem functions, the distributed system design, and the system users. The quality of the model's results are directly affected by the completeness and correctness of the system representation. The model should include all sources of instrument product workloads and all user service requests. It should include all the subsystems and the distributed computer and communication resources that have to be sized.

This metric is partially a measure of ECS functional traceability since it measures the completeness of the representation of the ECS functions and workloads. Further, it measures the correct representation of the design and its associated performance parameters such that components can be sized and performance requirements can be traced (see traceability metric below).

5.1.2.2 Testability Metric

The testability metric is defined in terms of the methods used for validation of the output results. The results from the model should eventually be validated by testing. Metrics should be defined in the model such that corresponding empirical measurements can be made. Further, the metrics should have statistical measures of variability computed in addition to mean values such as variances, percentile distributions, and confidence intervals. These statistics provide bounds for the metrics (i.e., the possible expected ranges for the metrics when they are measured).

Short of full system testing, there is additional validation analysis that can be performed on the model results. Prior to any actual measurement, the model results can be validated by analysis techniques (e.g., queueing network theory). When results are available from even limited measurements, such as benchmarks, more confidence can be gained in the model's results by calibrating the model's inputs based on measured parameter values, and by comparing the model's output to performance metrics measured for the limited benchmark cases. When the full system is tested, the same process can be performed on a larger scale for the model. Conversely, the validated model can be used to guide the testers by focusing the test cases on the critical cases, and also during the operational phase of the system for capacity planning purposes.

5.1.2.3 Traceability Metric

The Traceability metric is defined in terms of the ability of the model to assess design performance and track performance requirements. The statistics collected in the model should provide the data to make these assessments. The computer and communication resources have to be sized so that the performance requirements (e.g., response time) can be achieved. It is essential not only to make the correspondence of the performance requirements to the model components but also to be able to determine what the performance drivers are that impact the requirements compliance.

5.1.2.4 Flexibility Metric

The Flexibility metric is defined in terms of the model's ability to evolve to incorporate modifications. The model may need to be modified to examine deficiencies in its representation of the preliminary design (e.g., functions, design resources, or workloads missing from the model). It will need to be modified to incorporate additional levels of detail for the detailed design (e.g., additional details for the hardware, software, and communications networks) to evaluate design options and to demonstrate the performance requirements compliance at CDR. It may also need to be modified to address changes (proposed or actual) in requirements for new functions, new users , more frequent use, or different scenarios of use (i.e., evolution in how the users use the system).

The flexibility of the model to accommodate changes is probably the major factor in determining its long term usefulness as a performance analysis tool for the ECS program.

5.2 Constraints Affecting the Analysis

The following constraints have prevented a complete analysis of the performance model at this time:

- Limited time;
- Limited documentation; and
- Unavailability of model results and execution conditions.

The performance model was delivered by HAIS on January 25, 1995 This version of the performance model does not contain the complete functionality planned by HAIS or the latest version of the data provided by the AHWGP. This latest version is actually two separate models that are not integrated. Neither of the models was provided in executable form, and neither has a full set of input parameter data. We received an earlier version of the model on December 15, 1994. However, this preliminary model has limited functionality and only one set of instrument parameter values. Further, many substantial changes were made to this initial version. Also, at the time of this analysis, there was no documentation available on the HAIS SDPS preliminary design. The only ECS documentation available from HAIS is the System Design Specification for SDR. Therefore, there has been very little time and information for analysis of the latest version of the model.

HAIS has not developed documentation on the model and has not provided walkthroughs or other detailed descriptions others than what has been shown and provided at the AHWGP meetings. The BONeS simulation tool provides the capability to annotate the model modules. HAIS has not documented many of the modules (of which there are approximately 3,000). In general, the information provided represented a very limited set of what is needed for a complete understanding of the model.

We have received flow diagrams depicting data ingest and production processing for ASTER, MODIS, LIS, and MISR, and high level subsystem functional flow diagrams for the Data Handling, Processing, Scheduler, Ingest, and Distribution Subsystems. Many changes have been made in the BONeS model implementations, however updated diagrams have not been provided.

The only model parameter data provided for the "push" workloads is for the ASTER instrument. For the "pull" side, data has only been provided for the transaction request parameters (i.e., arrival rates, service types and functions, and initial DAAC location).

Since no model results were delivered with the model and neither model is in executable form, we have not been able to evaluate the model results.

As a result of these limitations, we have only done a preliminary evaluation of the model at this time. However, we have established a thorough criteria and methodology for evaluating the model. Appendix C contains many performance modeling tables that are incomplete in this report. This information can serve as the basis of a more comprehensive evaluation when a complete model is available.

5.3 Results

5.3.1 Discussion of Results

The results of the Performance Model evaluation are discussed below and are organized by the four evaluation areas: system design representation; model input parameters; model output statistics; and model structure. The model evaluation is based on the following sources: the User Scenario Functional Analysis and associated User Model data; data from the AHWGP and the Algorithm Theoretical Basis Documents (ATBDs); the System Design Specification (SDS) from SDR; discussions and briefing information from HAIS on what is in the model; and preliminary versions of the Performance Model. The tables and descriptions in this section and in Appendix C layout the complete evaluation of the Performance Model. The complete evaluation cannot be performed at this time due to the constraints given in Section 5.2. The evaluation items are based on current documentation and preliminary versions of the Performance Model and could change based on the final version of the model and documentation on the model and preliminary design.

5.3.1.1 System Design Representation

For this evaluation HAIS has provided two separate models, several input data files, and various flow diagrams, spreadsheets and parameter descriptions. However, there is very little consistency or integration among these different sources of data. The two models provided implement: 1) "push" workload generation and execution flows for AM-1 and TRMM instruments plus a partial representation of "pull" workload generation and execution flows for three service types (referred to as the Version 1 model), and 2) a separate "push" workload generator with 15 service types (referred to as the stand-alone "pull" model). If no distinction is made in the text, the reference is to the Version 1 model since it represents the most complete version.

Descriptions of the subsystems in the SDS define what architectural components should be in the model. Ideally, documentation of the preliminary design should provide another primary source for determining what design components should be contained in the model. However, since no design documentation was available at the time of our analysis, we used the preliminary versions of the model as a secondary source from which we derived design details. We have not constrained the evaluation with respect to what design components should be in the model by what is actually in the model; we have used it (because it is the only source available now) as a guideline but supplementing it with experience in modeling distributed systems.

The detailed set of exhibits shown in Appendix C will provide the basis for the evaluation of the model's completeness and correctness, with respect to the ECS design representation, when the final version of the model and associated documentation become available. The exhibits are TBD for this release of the TAR. Exhibit 5-1 contains a list of those Appendix C exhibits and a brief description of the contents of the exhibits.

Exhibit(s)	Title	Description
Exhibit C-1	Subsystem Sites	Subsystems at DAAC Sites
Exhibit C-2	Subsystem Services	Service Classes by Subsystem
Exhibits C-3,...,C-9	Subsystem Activities/Processes	Input and Output Activities by Subsystem and DAAC Site
Exhibits C-10,...,C-37	Data Files	Data Files by Instrument and DAAC Site
Exhibits C-38,...,C-65	Production Processes	Production Processes by Instrument and DAAC Site
Exhibit C-66	User Workload	User Services - Process(es) executed
Exhibit C-67	Production Workload	Production Processing- Process(es) executed
Exhibit C-68	System Resources	System Resources by DAAC Site
Exhibits C-69, C-70	System Overheads	Overhead by Workload and Overhead by Resource
Exhibit C-71	Process Input/Output Files	Input and Output File Requirements by Process

EXHIBIT 5-1: Exhibits Of Model Evaluation Results

The results of the evaluation of the preliminary versions of the model and the associated data are discussed below.

All of the subsystems described in the SDS appear to be included (to varying degrees of detail), with the exception of the Client Subsystem, in either the Version 1 integrated model or the stand-alone "pull" model. The subsystem services are represented by processing and/or storage service time delays computed from service demand and service capacity parameters. Network file transfers and user requests are represented by transfer delays also computed from service demand and service capacity parameters. A mapping of the subsystem representations is given below:

Data Server: Represented by the Data Handler and User Processing modules.

Ingest: Represented by the Ingest module.

Interoperability: Represented by the inter- and intra-site network links (resources); it is unclear whether the Advertising and Subscription services of the Interoperability Subsystem are represented.

Planning: 8-10 static rules are implemented in the Event Driven Scheduler module.

Data Processing: Represented by the Processing and User Processing modules.

Data Management: Represented in the Data Handler and Distribution modules - the stand-alone Pull model includes DIM, LIM, Data Dictionary, and Advertising Agent function and 12 more user services not included in the Version 1 model.

Client: There does not appear to be any processing and I/O representation of the Client services; however, the Pull Generator module represents the Client transaction sources.

The distributed DAAC sites are represented by computer resource arrays (see discussion of resources in Section 5.3.1.4 below) in the subsystem modules. Each DAAC site is assumed to contain all of the subsystems, and each subsystem is assumed to have its own set of computer resources. The model currently allows a maximum of 14 sites. The sites identified in the model input are: ASF, EDC, GSFC, JPL, LaRC, MSFC, and NSIDC. No details of the inter-DAAC wide area network (WAN) or the intra-DAAC local area networks (LANs) are represented other than the links (e.g., routers, communication devices, or protocol functions are not represented).

For the “push” workloads, data files and production processes are included for AM-1 (i.e., ASTER, CERES, MISR, MOPITT, and MODIS) and TRMM (i.e., CERES, LIS, GV, PR, TMI, and VIRS). Logic is included to represent network and physical media ingest and production processing and data storage for the instruments listed above plus two other categories: NMC and a catch-all category called “other.” The TRMM and AM-1 workloads included in the model inputs account for over 99% of data volume and processing requirements of Priority 1 requirements (from SPSO database numbers) and about 39% of data and processing requirements of the SPSO grand total. The SPSO numbers do not include V0 and reprocessing loads.

The model does not include loads for V0 migration, reprocessing, or platforms beyond TRMM and AM-1. Specifically not included are the following sources of data: ACRIM, AIRS, AMSU, AVHRR, CERES PM-1, CII, COLOR, DAO, DORIS, ESOP, ETM, GLAS, HIRDLS, MHS, MIMR, MLS, MODIS PM-1, SAGE III, SOLTICE, SSALT, SWS, SeaWiFS, TES, and TMR.

The “pull” model is in various states of detail and implementation. The most complete model (i.e., Version 1) that we have received has a partial implementation of user service request generation and execution flow through the subsystem modules. A revised “paper model” shows “pull” workload flow through a set of user services through one or more DAACs. HAIS has partially implemented this paper model. This separate model is just a “pull” workload generator with a larger set of user services and representation of Data Management processing. This enhancement has not been integrated with the other model at this point.

The stand-alone "pull" workload generator model implements 15 user service types. These service types should be mapped onto the list of 49 services identified in the User Scenario Functional Analysis (see Exhibit C-66) to evaluate completeness. For each user service type in the model, there are separate processes representing the Data Management functions, Advertising Agent (AA), Distributed Information Manager (DIM), Local Information Manager (LIM), Data Dictionary (DD), and the database management system (DBMS) function of the Data Server (DS). With probability p_i the execution flow goes to one of the five processes for the particular user service request as shown in Exhibit 5-2. HAIS has made estimates for the p_i's. Although not implemented in the "pull" workload generator, the paper model shows one or more flows from the DIM to the AA and LIM and from the LIM to the DBMS.

User Service Types	AA	DD	DIM	LIM	DBMS
Simple Search, 1 Site	p_1	p_2	p_3	p_4	p_5
Simple Search, M Sites					
Match-Up Search, 1 Site					
Match-Up Search, M Sites					
Coincident Search, 1 Site					
Coincident Search, M Sites					
Archive (Insert)					
Ingest					
Inspect (Browse)					
Produce					
Exchange					
Standing Order (Subscription)					
Manipulate (Subset)					
Modify					
Acquire (Exchange)					

EXHIBIT 5-2: User Service Types And Processes

The stand-alone "pull" model samples from several distributions in implementing "pull" transaction execution flows. These are described below:

- Diurnal Transaction Profile Distribution: determines service request frequency by time of day;
- DAAC Distribution: determines the source DAAC where the service request occurs;
- Service Type Distribution: determines which type of service is being requested;
- Service Function Distribution: determines which service function is executed for the service request (i.e., AA, DD, LIM, DIM, or DS); and
- Multi-site Distribution: determines the multiple DAAC sites for requests not satisfied at the source DAAC (the logic and input data for this part are not implemented in either version provided).

The Performance Model represents computer and network resources for the subsystems at each DAAC and for the CSMS network links. The resource types for the computers consists of processors, disks, read/write heads, robots, and input/put channels. Memory is assumed to be unconstrained and is not included as a resource. The network resources are the links between the DAAC sites and within the sites between the subsystem computers. EXHIBIT 5-3 shows the computer resource types represented in the major model modules. The network resources are included where transactions are routed between subsystems within a single site or between sites.

Model Module	Processors	Disks	I/O Channels	Robots	Read/Write Heads
Distribution		X	X		X
Data Handler		X	X	X	X
Ingest		X	X	X	X
Processing	X	X	X		
External		X	X		

EXHIBIT 5-3: Computer Resources Types

The resources are modeled as multi-queue, multi-server service centers with first-come, first-served (FCFS) preemptive priority service discipline. The assumption is that a resource server locks out all other lower or equal priority requests until the service requirements of the current transaction have been met . This assumption is not realistic given the nature of time-slicing operating system schedulers and network protocol fragmentation-reassembly algorithms. Direct overheads (i.e., serial delays for transactions) for functions like protocol, user interface, and file transfer software are not included. For user "pull" searches, the paper model shows database management software for the Data Server Subsystem, but this has not been implemented. However, no file management processing overhead is included for "push" workloads.

5.3.1.2 Model Input Parameters

Performance Model parameters may be characterized into workload, process, file, resource, and topology parameters. Workloads have parameters that provide their execution flow through the system resources, frequency of occurrence, and epoch(s) of activity. Processes should be parameterized by their execution site, resources used, resource service demands, and their input and output requirements. Files should be parameterized by their location, size, storage retention status, and specialized parameters relating data segmentation information. System resources should be parameterized by location, capacity, quantity, service and queueing disciplines, and indirect overhead. Finally, topology should be parameterized by site locations and connectivities within the site resources and among the sites (i.e., the local and wide area networks). The minimum set of specific parameter categories that should be included in the ECS Performance Model for the five parameter groups are summarized in Exhibit 5-4. The results of the evaluation of the Performance Model parameters are discussed below. The evaluation is mostly TBD at this time.

Parameter Group	Parameter Categories

Workloads	Category, Name/Id, Arrival Rate, Source, Destination, Epoch, Process(es), WAN Links, Link Service Demand
Processes	Category, Name/Id, Site/Component, Site Resources, Resource Service Demands, Input File(s), Output File(s)
Files	Category, Name/Id, Instrument, Size, Site Resources, Temporal Coverage, Archive Site, Storage Retention
Resources	Category, Name/Id, Site/WAN Link, Quantity, Background Overhead, Capacity, Service Discipline
Topology	Intra-site resource connections (resource ids), Inter-site connections (site ids)

EXHIBIT 5-4: Minimum Model Parameter Categories

In general the Performance Model contains the above set of parameters. The “push” workload parameters are more mature than those for the “pull” workloads. However, we have only seen a stand-alone model of the “pull” workload generator at this point. We expect the mature version of the full “pull” model to be comparable to the “push” model in terms of it parameter set.

One of the parameter areas identified above is overhead. In some cases in a Performance Model it is reasonable to represent some aspects of resource demand implicitly as an overhead to the resource capacity. The justification is that an explicit representation of the overhead may require a substantial implementation effort with only a marginal increase in accuracy of performance predictions. This approach requires experience to use; should be used judiciously; and have a realistic basis. In the current version of the model we have identified three uses of overhead factors:

- CPU efficiency - represents the level of processing efficiency possible as a percentage of the vendor’s claim; 25% is the current default value;
- Network link - for both inter- and intra-site networks, the percentage of unusable bandwidth due to protocol data and operation; the current default value is 0; and
- Overlap between CPU and disk service times- represents the degree of parallel/serial operations of the devices; 0 represents completely serial, 1 represents completely parallel; the current default value is 1.

The first two factors above are commonly used to represent background overhead (i.e., resource consumption that is not on the serial delay path of a transaction but contribute to transaction queueing delay). We have no indication at this point what the 25% CPU efficiency factor incorporates as overhead. It was a supplied by the Project Office for use in de-rating CPU capacity. The factor can be used to represent different mixes of instruction executed, operating system functions, and other system management delays such as swapping, paging and interrupt handling. None of these overhead activities are represented explicitly in the model.

In the current versions of the model, the network overhead value is zero. Overhead should be included for protocol headers and trailers, control packets (e.g., flow control and acknowledgments), and multi-user access methods where applicable (e.g., CSMA/CD or token passing). This kind of information can be derived from information contained in protocol specifications.

The third overhead factor listed above is suspect in terms of producing accurate results. While it may capture the total serial service delay that a single transaction experiences between the two types of resources, it will lead to error in determining the queueing delay experienced by all transactions that pass through the resources. Another problem with using this approach is in determining a realistic basis for the estimate of the factor. A reasonable estimate will have to be derived from empirical data from a prototype or an actual production environment. While sensitivity analysis can be performed with a range of values for the factor, it is still necessary to know where within the range a nominal value will fall for the given set of applications. Finally, there will be a problem in accurately separating resource utilization for the CPUs and disks. This will make sizing those resources difficult using the results from the model. A separate sizing approach would probably be required.

5.3.1.3 Model Output Statistics and Performance Metrics

The basis for evaluating the Performance Model in the area of output statistics is the EOSDIS Functional and Performance Requirements and commonly used performance metrics in evaluating distributed systems. The results of the evaluation are summarized below. Exhibit C-72, Resource Metrics, and Exhibit C-73, Performance Requirements Compliance, in Appendix C provide the basis for evaluation of the final version of the model and its results.

Since no evaluation conditions nor results were provided, we cannot assess those characteristics of the model. We can, however, assess what performance statistics the model collects.

The model has numerous statistics collection probes defined. One class of BONeS probes produces resource utilization and number of resources in simultaneous use. Examples of these probes can be seen in the top level BONeS diagram in Exhibit 5-6. These are valuable in assessing the design in a limited way but do not allow evaluation of many of the performance requirements (e.g., response times).

EXHIBIT 5-6 BONeS Block Diagram

It should be noted that the steady state resource utilization-type statistics described above can be computed by a simple formula based on arrival rate, resource service demand, and resource capacity, and do not actually require the execution of the simulation model to quantify. However, the distribution of utilization over time cannot be produced practically without the simulation model. The utilization-type probes that have been defined in the model are listed in Exhibit 5-7.

Resource	Ingest	Data Handler	Distribution	Processing	Interoper-ability	External
Read/Write Heads	X	X	X			
Robots	X	X				
Processors				X		
Disks	X	X	X	X		X
I/O Channel	X	X	X	X		
Inter-site Links					X	
Intra-site Links					X	

EXHIBIT 5-7: Utilization-type Statistics Probes

The other class of BONeS probes collects system statistics such as response time and throughput across multiple system modules. These probes are defined for the BONeS model as a part of the simulation execution configuration conditions. Since we do have access to an executable version of the model at the present time, we cannot assess the inclusion of this type of probe. However, based on verbal information provided by HAIS, we know that there are probes defined for simulation level (e.g., we have seen throughput results provided by the HAIS modeling team).
Based on verbal information provided by HAIS, the model has been executed for individual TRMM and AM-1 instrument loads but not for combined “push” loads. Also, since the “pull” workload portions of the model have not yet been integrated, those workloads have only been executed in the stand-alone “pull” model. We have no indication whether resource utilization and other performance metrics have been assessed for the “pull” workloads.

5.3.1.4 Model Structure

In general, the ECS Performance Model is well structured. Subsystems are represented separately in modules, and resources within subsystem modules are represented parametrically. These features will aid in minimizing modification effort. Many workload and service demand parameters for the modules are read from files or are provided via screen input. Consequently, many changes can be made simply by data input. There are areas, discussed below, where changes to the model will ripple through multiple portions of the model.

Exhibit 5-6 shows the actual BONeS block diagram (from the BONeS Block Editor) for the highest level of the Performance Model. The BONeS system modules at the top of the diagram represent the major SDPS subsystems, the "push" and "pull" workloads, and other necessary modeling functions. Each module in the diagram has many more block diagrams underneath it arranged hierarchically. There are over 3,000 total modules in the model. The diagram shows a probe module (at the top of the diagram) that implements user-defined statistics collection functions. The diagram also shows model parameters (represented by the letter 'P'), system resources (represented by the letter 'R'), and model variables that are shared across multiple blocks, called memory (represented by the letter 'M'). Model parameters, resources, and memory can be defined for each block level. Parameter values are provided from file input or screen input.

The BONeS system modules implement the functions of the ECS Subsystems and "push" and "pull" workloads. Subsystems for DAAC sites, site resources, interconnecting network links between the sites, and users from different locations are represented in these modules. The system resources, which are actually arrays in the model software, represent the processors, disks, network links, etc. at the DAAC sites or between the sites. These resource arrays are indexed by site and resource type. Process and file parameters, read from files in some of the system modules, populate data structures, or memory. Some of elements of the data structures contain values for the resource array indices. The values are used by routing functions implemented in the system modules. Workload execution flows are implemented by the data structures being routed from one system module to another. Service demand values for processes, files, and network data transfers (i.e., messages or files) are also read from files and passed as memory variables to the appropriate system modules in the simulated transaction flow. Parameters such as resource capacities are provided by values and expressions read from screen input. These are combined with the resource service demand parameters in system modules to compute resource service times for transactions.

Existing representations of system workload, architecture, and design can be modified usually by changing parameter values; however, additions or deletions in most cases will require modifications that impact more than one major module. For example, existing workloads can be changed without affecting other modules by modifying file inputs, and processes can be move to different sites by the same method. Quantities of existing resources within an existing pool can be changed by modifying resource parameters with no impact to other parts of the model. Statistical probes can be added with little or no perturbation to other modules. However, new workloads, subsystem functionality, and resource pools cannot be added without adding modules or resource pools and making appropriate changes in other modules (e.g., mapping workload execution flows onto resources).

5.3.2 Identified Problems

The most immediate problems are: the model does not include all workloads for the TRMM and AM-1 timeframe; it is not yet fully integrated; and a performance analysis of all the "push" and "pull" workloads has not been performed.

A second set of problems relates to a more accurate representation of system resources and system overheads. This is important for accurately predicting performance (for assessing requirements compliance) and for sizing the resources (as input to the cost model). The three main problems are :

- No direct processor resources are represented for subsystems other than Processing;
- Network and computer system overhead functions have mostly been ignored; and
- Service queueing disciplines are all assumed to be first-come, first-served.

The consequence of the first two problems are that delay and processor utilization will be underestimated, and unless other techniques supplement the simulation model for sizing the processors, cost may be underestimated. The consequence of the last problem is that delays will be inaccurately overestimated. The impact of this could be the addition of resources (to meet timing requirements) which could translate into higher system cost. These additional resources might not be necessary to meet the performance requirements. Finally, we have no indication that appropriate statistics probes have been defined. This is necessary to assess requirements compliance.

5.3.3 Potential Issues

The major potential issue for the CDR timeframe is memory sizing. The current model does not represent memory. It is not necessary to model memory usage in the system model at PDR. However, once detailed software designs are developed, memory requirements should be developed. This can be more appropriately addressed with subsystem models.

The absence of the Client Subsystem in the Performance Model is another potential issue because it is a component of delay for user response time requirements. This is not an issue for The TRMM Release because the V0 IMS software is expected to be used in place of the Client software. However, there should be an assessment of delay and utilization of the Client software for the timeframe beyond the TRMM Release.

5.4 Conclusions

5.4.1 Engineering Quality

The engineering quality of the Performance Model is somewhat limited because the model is incomplete and not fully integrated. However, the level of detail is reasonable to assess the preliminary design once enhancements are made and integrated, and all workloads are considered in the performance analysis.

Although the model is not integrated at this point and no analysis has been performed with all of the “push” and “pull” workloads, the model appears to have the necessary workload capabilities to assess the TRMM and AM-1 releases with the exception of reprocessing and V0 migration loads. These missing loads are significant and will affect performance and cost.

The model assesses processing requirements only for the Data Processing Subsystem. The processing demand for the Ingest data input (e.g., for interrupt handling) is bound to be significant because of the high data rate even though science processing algorithms are not being executed. The processing delays for Ingest and the other Subsystems are serial delays and contribute to processor loading. They need to be included to size the processors for costing purposes and to assess the performance requirements compliance.

Adequate estimation of system overhead performance parameters has not been made at this point. Network protocols like TCP/IP, among others, and operating system functions, will impact performance and cost.

The exclusion of the Client Subsystem is a secondary concern for PDR; however, the processing within Client Subsystems is part of the performance requirements timeline for user service response times and provides the basis for sizing the Client hardware. Processing and network service demand requirements for Client graphic user interface software are typically significant. Even though HAIS is not providing the Client hardware, requirements for a minimum configuration to host the Client software should be developed.

Another potential performance and cost issue is computer memory. The model does not address memory usage. This is reasonable for PDR. However, because of the large data and processing memory requirements it should be considered for CDR.

5.4.2 Testability

The testability of the Performance Model's results at this point is nil because the model is at a state of very limited maturity in this area.

5.4.3 Traceability

The traceability of performance requirements in Performance Model at this point is nil because the model is at a state of very limited maturity in this area.

5.4.4 User Satisfaction

The ECS Performance Model and its results can be used to great advantage by the HAIS designers, the NASA Project Office, and the instrument teams. The model has apparently already been used to a limited extent by the HAIS designers. The modeling team has reported that design decisions have been made based on modeling results. As HAIS refines the design to lower levels of detail, they will be able to base design decisions on model results of the design alternatives. The instrument teams will be able to develop strategies for reprocessing and refining science processing parameters based on analyzing resource consumption results from the model. The project team will be able to assess requirements compliance, cost implications, and design robustness to requirements changes. However, because the model is very large and complex, and because of the performance modeling expertise required, it will probably not be directly used extensively. Users other than the HAIS and IV&V modeling teams will require extensive training to use it effectively.

For all of the users of the model, it is extremely important that the model's results can be trusted. Therefore, the validation of the model and its ability to produce results that can be used to assess the design is most critical.

5.4.5 Trends and Projections

For future releases of the modeling TAR, this section will document what the track record of the model is (i.e., has the model become more useful since the last snap-shot). This evaluation represents the baseline point for future assessments.

The flexibility of the model to be modified easily to address deficiencies will be a driving factor in whether the model will become more useful in the future. The model appears to have sufficient flexibility to support modifications with reasonable level of effort.

5.5 Recommendations

5.5.1 Areas Requiring Further Analysis

In this document we have defined the evaluation criteria and methodology for evaluating the Performance Model. Because an executable, integrated Performance Model has not been provided to us, we have only been able to perform a preliminary analysis of an incomplete model. When the complete, integrated Performance Model is available, a thorough evaluation of the model and results generated by the model using the methodology defined in this report should be performed. Some of the tables in Appendix C are expected to be refined and expanded in the process of the detailed evaluation. As a part of that evaluation, available documentation on the preliminary design and the model should be examined, and a detailed performance analysis should be performed using the model.

5.5.2 Solutions to Important Problems

The HAIS modeling team appears to be on the right track for solving the problem of integrating the separate models into a single model. The model will not be available to us to support PDR analyses.

The problems of adding the V0 and reprocessing workloads will require a moderate amount of modifications to the model. The model may be able to accommodate the reprocessing loads simply by increasing the direct processing arrival rates or with simple modifications (e.g., implementing a new workload generator that uses the same processing functions as the direct processing workloads). Implementing the V0 workload processing will probably require more effort (e.g., adding new production processing logic in addition to a new workload generator).

The problem of representing resource service disciplines as first-come, first-served (FCFS) can be corrected with minimal effort. BONeS provides the capability to model service disciplines as round robin with user-specified time slices. For example, to model the packet fragmentation for the network file transfers, the time-slice is computed from packet size and the link bandwidth adjusted for protocol overhead. It will probably be necessary to make other changes to reflect the nature of the pipelining of data between the disks across the network. This is a very straight-forward approach to correcting the inaccuracy of using FCFS.

The problem of adding processors, processes, and service demands to account for processing on Subsystems other than Data Processing will require a significant modifications to the model. Process service demand parameters must be defined in the model and estimates for service demand must be made. Processor resources must be defined and logic modified within existing Subsystem modules to map the workload flows through the processor resources. System overheads can be determined from protocol specification, empirical data, and performance-oriented journals (e.g., Computer Measurement Group Proceedings) and periodicals.

Defining additional statistics probes (to assess performance requirements compliance) is conceptually easy given the capabilities of BONeS in this area. These can be defined with minimal effort once an integrated and executable model is available. In some cases the model may need to have statistics variables defined in the data structure and minimal logic added in a few modules to collect data to implement the probes.

5.5.3 Risk Management

To contain risk on performance requirements compliance of the ECS design and on accurate cost estimation, we recommend that the Project Office support the following activities:

- Continued model assessment throughout the lifecycle;
- Assess cost and performance impacts of problem areas identified;
- Conduct sensitivity analysis of key parameters (i.e., what-if analysis for the performance drivers and parameters where uncertainty exists); and
- Support instrument teams in refining their model inputs (e.g., making model runs and providing results to the teams); and

Assess memory requirements and performance impact of memory.

6. COST MODEL

The Cost Model is intended to estimate the resources required to develop and operate ECS architecture alternatives (as partially derived from the Performance Model) within schedule constraints. The Cost Model is currently implemented as a collection of three types of stand-alone cost estimation methods: Custom Software, Operations and Maintenance, and Commercial-Off-The-Shelf (COTS) Hardware and Software. The findings of an independent assessment of the cost modeling activities in each of these three areas, including analysis tasks performed, constraints affecting the analysis, tools and databases utilized, and analysis results, conclusions, and recommendations, are presented in this section.

6.1 Analysis Tasks Performed

The cost modeling analysis entailed examination of the Custom Software, Operations and Maintenance, and COTS Hardware and Software cost estimation methods employed by the ECS contractor (i.e., HAIS). These three estimation methods correspond to the cost breakdown being used by HAIS and encompass the vast majority of the costs associated with developing the system. Because the modeling done by HAIS is different in each of the three areas, the evaluation for each area was somewhat different. Hence, the work done in analyzing each area is described in separate subsections below.

In performing the analysis, a common set of evaluation criteria was used. Key evaluation areas include the following:

- Completeness
- Correctness
- Accuracy
- Technical integrity
  - traceability to requirements
  - engineering quality
  - testability
- User satisfaction
  - support for the engineering process
  - implementation

The detailed meaning of the evaluation criteria is described in Appendix D. It was not possible to analyze all of these areas for every modeling area. Therefore, the items analyzed for each modeling area are reported in the separate subsections below. Second, some of the evaluation criteria have somewhat different meanings when considered in the context of a specific modeling area. Those specific meanings are also documented in the relevant subsections.

6.1.1 COTS HW and SW Estimation

Three models have been utilized by HAIS in the estimation of COTS hardware and software costs: the COTS cost estimation model; the COTS procurement model (a.k.a. the Bill-of-

Materials procurement cost model); and the distribution cost sensitivities model. The COTS cost estimation model and the COTS procurement model were examined in this work.

The parameters that tend to drive cost, and are therefore strongly related to accuracy, in the area of COTS hardware and software include numbers of components, the cost per component, and cost trends. Regarding number of components, the key issue addressed was how the numbers were derived. This in itself is a complex issue. There was no attempt in this analysis to duplicate the performance modeling analysis reported in Section 5. Rather, the goal was to determine whether performance modeling was used or not, and if not, what techniques were used. For component cost, the obvious key issue is whether the costs used for specific components are similar to those available in the marketplace. Given the historical increase in performance and reduction in cost as a function of time, and the fact that any major system such as EOSDIS is always built over a period of time, the key issue with cost trends is whether that historical performance improvement, or cost reduction, has been accounted for in estimating costs.

Regarding user satisfaction and the COTS cost estimation, the key issue is whether the model supports its primary purpose, i.e., trade-off analysis. Speed and ease of use are key to achieving that purpose. In contrast, regarding the COTS procurement model, the key issue is its ability to produce accurate costs. For this model, speed and ease of use are less of an issue. Attention must be paid to details regarding the specific components used, their numbers, and their costs.

In order to evaluate these issues a variety of methods was used. The initial plan was to become familiar with the models, their purposes, and the general way they were implemented, and then to obtain the models and examine them directly. When it became clear that the models were not going to be made available, the analysis became more focused on learning about the models and performing the evaluation on the basis of that knowledge. The specific methods used are described in Appendix D.

6.1.2 Custom SW Estimation

Unlike the COTS hardware and software estimation, custom software estimation follows a single approach. However, this approach is comprised of several steps. Therefore, each step had to be evaluated separately.

Completeness and correctness could not be evaluated due to a lack of information. The majority of the effort was focused on determining what the steps in the process are and how they are being performed. It was determined early in the evaluation process that a somewhat new method was being used to estimate software size. Therefore, a key aspect of the analysis immediately became whether that method was valid or not. Some of the specific tasks performed included:

- Determining what methods exist for estimating software size/effort given an object-oriented design (OOD) process and the degree to which they have been tested;
- Searching for an established relationship between source lines of code (SLOC) and (OOD) entities; and

- Empirically determining the relationship between SLOC and OOD entities such as objects and methods in existing C++ software.

The analysis was performed primarily by interviewing the model developer and analyzing the stated approach, as reported for COTS hardware and software estimation above. Since, the estimation approach being used is somewhat new, the analysis also involved querying the literature and engineering community for existence/familiarity with stated approach and for recommended strategies. The specific methods used are described in Appendix D.

6.1.3 Operational Cost Estimation

Operational costs are driven by personnel costs. Therefore, the key issues regarding estimation of operational costs center around how personnel costs are estimated. Completeness in this context refers to whether all the required types of staff are accounted for. Correctness refers to whether the types of personnel included are all required. Accuracy involves analysis of the rates utilized and whether the upward trend of personnel cost as a function of time is accurately represented.

The analysis was performed primarily by interviewing the model developer and analyzing the stated approach. The staff allocations performed for SDR were provided to aid understanding of the organization of staff functions. However, the numbers of staff reflected in this material were characterized by HAIS as out-of-date and invalid, and therefore, were not analyzed.

The average cost of a man-year was computed to compare with the figure(s) being used by HAIS. This was done by assuming a typical mix of labor categories and costs per labor category. The average cost resulting from these assumptions was then computed and compared to the stated value.

6.2 Constraints Affecting the Analysis

Overall, the non-availability of models and cost information severely limited the degree of analysis that could be performed in at least two ways. First, although the HAIS staff are knowledgeable and experienced, and their comments and described approaches reasonable, it is strictly impossible to verify the accuracy of those comments without access to the models or the underlying cost information. The only other approach is to perform independent estimates, which is outside the scope of this task. Second, without concrete information, the analysis is limited to examining the validity of the stated approaches; no analysis of cost accuracy is possible. The specific constraints encountered within each modeling area are documented in the following subsections.

6.2.1 COTS HW and SW Estimation

Within this modeling area the primary constraint was the non-availability of models and supporting detailed cost information. In general, the model developer made himself available and took the time to review write-ups and answer questions. As a result, a fairly clear and detailed picture of the modeling done in this area was obtained. However, as stated above, none of that discussion could be validated without access to the models themselves.

6.2.2 Custom SW Estimation

Constraints in the custom software estimation area were more numerous. First, it was not possible to have the thorough type of interchange regarding this modeling area that was achieved for COTS hardware and software. In addition, the very nature of this type of estimation is a constraint in itself. At this stage in the system life cycle, there is very little detail in the design of the software. As a result, producing counts of software entities is a subjective process, and validation of that process is more complex than it would be in later stages of the life cycle. Moreover, the fact that the design is object-oriented also adds some complexity and subjectivity to the estimation process. This is a result of the fact that there is no widespread experience with estimating size on the basis of counts of object-oriented design entities, and no thoroughly validated and calibrated procedures.

6.2.3 Operational Cost Estimation

Constraints affecting the analysis of operational cost estimation are primarily related to the non-automation of the approach. In many cases, the numbers of personnel are estimated in a purely manual way, based upon experience, and not upon established metrics. Although this may be typical for this type of cost estimation, it causes the analysis to be more difficult and time-consuming.

6.3 Results

It should be noted that the results reported here assume that the information provided in interviews with the ECS contractor is accurate. Attempts were made to have HAIS validate the interview notes. While this happened in some cases it did not happen in all cases. Further, very little hard or soft copy data was made available to cross check the information gleaned from the interview process.

6.3.1 Discussion of Results

6.3.1.1 COTS HW and SW Estimation

The COTS cost estimation model and the COTS procurement cost model have been documented by HAIS in separate documents. Therefore, only the points not covered there or important to this analysis will be reported below. The COTS procurement cost model will be used to develop the hardware and software cost for PDR. However, this cost will only represent the TRMM and AM-1 releases. Costs across all releases will be computed at a later time.

6.3.1.1.1 COTS Cost Estimation Model - This model is a cost estimation tool that provides an estimated cost for COTS hardware and software, associated maintenance, and operations costs associated with the COTS hardware through the end of the contract. The model provides a decomposition of the system costs into element costs and the ability to estimate costs for processing hardware and software.

This model decomposes the system based upon the design submitted as part of the original HAIS proposal. It generally follows the decomposition represented in the F&PR, i.e. SDPS (PGS, DADS, IMS) , CSMS (SMC/LSM, ESN), etc. However, it decomposes the system a step further into "sub-elements". For DADS, this includes archive, ingest, and distribution. These sub-elements are shown on Figure 2.1.1-1 of the HAIS proposal.

This model provides the capability to estimate costs for processing hardware based upon the MFLOPS estimates (such as those from the SPSO database), and RAID disk and storage archive based upon the data volume estimates (also available from the SPSO database). The cost of each sub-element can be estimated by one of at least four ways. The method used for each sub-element is controlled by the operator. The four methods are described below:

a) Based on data volumes / MFLOPS from the SPSO database or the AHWGP.
b) Based on the following formula: technical baseline data volume / SOW data volume * cost from proposal for the element and sub-element in question.
c) Based on the following formula: technical baseline granule volume / SOW granule volume * cost from proposal for the element and sub-element in question.
d) Added directly from contract or Change Order #1. (Note that Change Order #1 cost additions were derived by costing the added configuration items using the COTS Procurement Cost Model).

In the case of methods b) and c), the technology which formed the basis for the contract cost is used as the basis of the estimate.

When costs are estimated for the primary components of the system, price performance curves are used to determine what the price of a particular item will be when it is purchased *n* years in the future. The starting point for processor capacity and price is the average price/MFLOP of the SGI Power Challenge and the DEC 7000/620. A comparison between the annual reduction in price per unit of performance utilized by HAIS in the COTS cost estimation model versus those estimated from representative data for the major components of the system is shown in Exhibit 6-1. The data points and formula which were used in computing the estimated price / performance change per year are presented and described in Appendix D.

Technology	Historical Increase in Capacity	Estimated Price/ Performance Change	Utilized Price/ Performance Change
CPU (Price/MFLOP)	133% / year	0.34-0.40 / year	0.21 / year
Disk (Price / GB)	100+% / year	0.35-0.44 / year	0.28 / year
Archive (Price/TB)	100+% /year	0.32-0.34 / year	0.05 / year thru 99 0.32 / year 99 - 02

EXHIBIT 6-1: Estimated Versus Utilized Annual Reduction In Price Per Unit Performance

Computation of numbers of components required also included the use of a 25% efficiency factor. This is a contract requirement. This is obviously a major cost driver. More typically, a 50% efficiency factor is utilized. However, it would also be customary to add an uncertainty factor of 50% to any workload computations at this phase of the system life cycle. This was not done. Therefore, the 25% efficiency factor may be justified.

When the COTS cost estimation model computes hardware and software cost directly (rather than as a ratio of current capacity estimates to previous capacity estimates multiplied by previous cost estimates), it also uses standard multipliers to include cost for hardware maintenance and operation of the processors. The multiplier for hardware maintenance is 9.5%. Operations cost is estimated by assuming one operator per eight processors for day shifts and one operator per 16 processors for other shifts.

6.3.1.1.2 COTS Procurement Model - The COTS procurement cost model, also known as the Bill-of-Materials Procurement Cost Model, is a means of estimating COTS hardware and software costs given a bill-of-materials. Inputs are provided to this model in the form of code names for each item in the bill-of-materials, the required quantity, and the release for which the items must be procured. Given this set of inputs, the model equates the code name to a specific make and model of hardware or software, converts the release entry to a date, applies phasing, then applies a cost as a function of time curve. No information regarding the rates of decrease of price per unit performance used within this model was obtained. The output of the model is the total COTS hardware and software cost across the life of the project for the input bill-of materials. This model also has the capability to compute maintenance costs as a function of time.

The entire model, including the cost versus time curve(s), is provided by EDS (another Hughes subsidiary). The beginning cost in the cost versus time curves are obtained through a bidding process followed by a negotiation process. The implication is that they are very competitive, however, this could not be verified. Costs in the COTS Procurement cost model are updated whenever repricing activities (such as Change Order #1 updates) occur. At such times, EDS revalidates the vendor quotes and their pricing curves. This has occurred at least three times since the start of the contract (August, 1993, April, 1994, and for PDR).

Review of a sample output of this model (sans costs), corresponding to the negotiated baseline plus Change Order #1, determined that in general, the list is fairly detailed, including cables, media, etc. Software, however, is represented as "bundles", hence, there is little insight into what is contained there. According to HAIS, numbers of users are reflected in the software costs. In general, the example model output seems out of date. Given that this represents the original bill of materials included in the proposal, that may be understandable. Some examples of this are:

- There are several cases where large numbers (40 or more) of 780 MB, 5.25" form factor disks are being employed. It would seem that costs would be reduced if fewer 3.5" form factor disks of higher capacity were used. There does not seem to be a consistent pattern of using these where database functionality is predominant, which would be understandable (to increase the number of spindles, thereby increasing database performance).

- At least four different archive technologies are included for each site, including 3480 drives and 3480 cartridges (for archive). 3480 is an out-of-date technology with very high cost per TB. Further, it would seem preferable to use a single technology at each site, and perhaps system-wide.

6.3.1.1.3 Model Implementation - In the case of the COTS cost estimation model, the model has been implemented as several Excel spreadsheets. The COTS procurement model is implemented as a series of dBASE files (database files and program files).

Running the COTS cost estimation model is a manually intensive process. It is incumbent on the operator to account for all functions, sites, changes, etc., and insure that all relevant costs are included. Moreover, the spreadsheets used to implement the model are not linked. The operator must manually cut-and-paste outputs on one sheet into another.

6.3.1.2 Custom SW Estimation

6.3.1.2.1 Methods used by HAIS -The method used by HAIS to estimate software costs can be subdivided into three steps: 1) estimate software size, 2) estimate effort and schedule, and 3) estimate cost.

Software size was estimated by performing the following:

a) Count objects/classes, as contained in the System Design Specification.
b) Characterize each object/class as simple, average, or complex.
c) Estimate the numbers of operations per object/class.
d) Multiply the number of operations in each object/class by fixed numbers of SLOC based on the complexity characterization:

simple	100 SLOC / operation
average	150 SLOC / operation
complex	200 SLOC / operation

e) Sum the number of SLOC across all objects/classes.

f) Add a ten percent margin of safety.

Effort/schedule was estimated by performing the following:

a) Applying REVIC (a Hughes proprietary tool which is similar to COCOMO) to the SLOC estimate derived as described above to obtain effort/schedule.
b) When the schedule predicted by REVIC was longer than the time available, subsystems were split into two parts and assigned to separate teams. Some efficiency adjustments were made in these cases to account for integration issues.
c) The requirements of each release were examined to determine allocation of functionality to each release. In general, functionality was added as late as possible.
d) In some cases, a fundamentally different type of capability was needed from one release to the next, resulting in rework. Rework was accounted for.
e) Different levels of productivity were assumed for different types of code. Also, productivity was assumed to be lower than average in the early releases and higher than average in the later releases to account for the learning curve with C++.

Cost was estimated based on man-month estimates for each subsystem and a parameter indicating the percentage of senior staff. These estimates were then multiplied by fixed rates and then summed to achieve the total software cost.

This model was implemented using Excel and REVIC. The size estimation described above was implemented as Excel spreadsheets. Effort/schedule are estimated using REVIC. Cost estimation given the effort and schedule determined by REVIC is performed using Excel. The transfer of data between the two tools is done manually.

6.3.1.2.2 Methods Used Elsewhere **-** The results of efforts to empirically determine the multipliers to be used when estimating code size based on an object-oriented design are provided in Exhibit 6-2. The sources listed in this table are described in further detail in Appendix D. It should be noted that the NIH class library represents a set of code that has been carefully designed for reuse and which is fairly mature in that area. In contrast, the FAST PPS and SAMPEX CMS libraries were developed for operational missions with less emphasis on reuse and more emphasis on getting the job done. Hence, the NIH Class Library tends to have more methods per class and many more SLOC per method. The GSFC developed code is more likely to be representative of the code that will be produced for EOSDIS. It should also be noted that these repositories represent software towards the end of a project lifecycle. Therefore, the ratios obtained must be scaled to correspond to the relationship between SLOC and methods at earlier phases in a project.

Source	SLOC / Method simple	SLOC / Method average	SLOC / Method complex
NIH Class Library, .c	25	43	82
FAST PPS	7	n/a	n/a
SAMPEX CMS	10	n/a	n/a

EXHIBIT 6-2: SLOC Per Method In Existing C++ Software Libraries

Questions were posted on the Internet news system to solicit information on software estimation techniques used for systems being developed under object-oriented approach. There were two primary responses. The first was from Greg Wenneson and John Connell of Sterling Software in support of the Software Engineering Process Group at NASA Ames. They have published a paper on the topic of software estimation in conjunction with object-oriented design. Their paper [31], validates the notion that a relationship between object-oriented designs and software effort can be derived. In particular, they produce a metric called object-oriented units which are similar to function points. Therefore, their approach is not directly comparable to the HAIS approach. The other useful piece of information provided in this paper is that the size/effort estimates produced in early stages of a project tend to be off by a factor of four. The second response merely indicated that Intel has also utilized a counts and multipliers approach to estimation of object-oriented software.

Reifer Consultants, Inc., who are experts in the area of software estimation, state unequivocally [30] that a size estimate can be considered complete when and only when:

- Two estimates have been done using different techniques,
- The two estimates have been compared and any major differences explained,
- The final estimate has been verified, and
- The baseline estimate is established and documented.

In other words, software estimation is a very imprecise science. When a lot of dollars are at stake it is very important to do the software estimation very carefully and consider multiple approaches.

6.3.1.3 Operational Cost Estimation

Operational cost estimation is performed primarily to support the bid process. Examples include the original proposal, Change Order #1, and Change Order #2. When a trade analysis is performed, a partial estimate of operations costs is computed and included in the cost estimate by the COTS cost estimation model.

Operational cost estimation includes costs in the following categories:

- Operations staff
- Maintenance staff

- Integrated Logistics Support
- Sustaining engineering
- Management
- Training

Staffing is estimated for each operational entity. The general process of estimating staffing levels is: 1) identify staffing requirements for each system function; 2) determine the size of each operational entity (from modeling results) and associated workload; 3) determine size of code; 4) estimate staff; and 5) allocate staff. Although there are areas in which the number of staff is estimated according to a mathematical formula, the process is not quantitative in general. Rather, staffing levels, allocation of staff to locations, and allocation of responsibilities to staff are decided more on the basis of experience and judgment.

Where quantitative measures are used, they are based on the Technical Baseline defined for PDR. Those areas for which quantitative measures are being used include:

- Number of lines of code maintainable by one person:
  18,000 - first year after release
  36,000 - second year after release
  48,000 - third year after release

- Time to handle a single media, including mounting, dismounting, etc.: 6 minutes. Rereading media to insure the data has been written correctly is not envisioned at this time. The steps included in this process are loading and unloading of media, pickup of shipping documentation, placing labels on media, Q/A of the shipping materials (insure that all the pieces and the correct pieces are together), and packing of media and shipping documentation. The six minutes has been reduced to five based on automation . Packing of media and labeling for shipping may be done by the shipping department of the local facility if it is cheaper to do so. However, the costs for these activities are currently being included.

- Cost for hardware maintenance: 8-10% of the hardware cost. The hardware configuration is the basis for maintenance costs. The configuration comes from system integration and planning (HAIS). Regarding maintenance, some level of on-site maintenance has been requested by GSFC, over and above vendor maintenance. The level of this support is TBD at this time.

- Effort will be expended to develop a quantitative measure for DBA activities based on information from the AHWGP.

- Number of operators is driven by the number of processors. Current thinking (subject to change) is that one person per 8 processors will be sufficient for daytime operations. One person per up to 40 processors is being considered for nighttime operations.

Regarding the effect of the level of automation on staffing estimates, the general guideline followed was that staffing stays constant while workload (productivity) increases due to automation. Two examples were discussed, data distribution and data production (planning and scheduling):

> **Data Distribution**. The TRMM Release is assumed to be very low automation. In this case, a multiplier of 10 minutes per piece of media is utilized. In the AM-1 Release, the use of bar-code labels and bar-code readers, as well as media autochangers is assumed. As a result the multiplier per piece of media reduces to 5 minutes.
>
> **Data Production**. The TRMM Release is assumed to be batch oriented (production is primarily operator controlled). In the AM-1 Release, the current thinking is that production becomes software controlled, therefore allowing the same staff to handle increased data volumes.

The general trend is to centralize staff functions in the larger operational entities while leaving at least one person at the smaller sites to provide each staff function. In many cases one person wears several hats at the smaller sites.

Currently, the assumption is that ECS will have its own dedicated staff. At some time, the possibility of using existing staff at the various locations to perform some of the duties will be investigated. However, the assumption of dedicated staff will carry through PDR. In addition, the staff being estimated contains all staff required to do the job regardless of whether they are government or contractor staff. According to HAIS, for costing purposes, $100k per man year, regardless of position, should be assumed.

Most of the staff sizing is done manually. The staffing levels/allocations are being maintained in a spreadsheet. Whether the quantitative measures mentioned above are embedded in the spreadsheet is TBD. Whether the computation of the resulting cost is implemented in the spreadsheet is also TBD. Costing may be done by a separate group based on staffing levels.

No responses were received from the Internet news system in reply to queries regarding the existence of commercial tools for operations cost estimation.

6.3.2 Identified Problems

In general, few aspects of the modeling activities were definitively determined to be serious problems. The lack of sufficient information provided prevents us from reaching firm conclusions. In addition, some of the established problems pose a somewhat minor impact. A large number of potential problems were discovered. These are discussed in Section 6.3.3, below.

6.3.2.1 COTS HW and SW Estimation

In the COTS hardware and software estimation area, three problems were identified regarding the COTS Cost Estimation Model. First, the model is overly manual, leading to a greater chance of errors. Second, the practice of multiplying old costs to get future costs leads to overly high estimates because the costs computed in this way are based on dated price/performance data. This is not envisioned to be a serious problem as long as 1) the model is used for trade analysis and not for life-cycle cost estimation, and 2) this method accounts for a minority of the costs being estimated. It is not clear that both conditions have been satisfied for estimates produced for PDR. The third problem identified regarding the COTS Cost Estimation Model is that, in general, most of the parameters used are very conservative. One example is the compression ratio assumption for archive of 1.5. Two independent studies have found that the minimum compression ratio achieved using lossless compression of satellite data is about 1.7. However, compression ratios as high as 5 were demonstrated with the average well in excess of 2. Other parameters demonstrate this conservative trend as well. A key set of parameters with respect to cost estimation are the price/performance rates of decrease. Since the values computed independently in this analysis were not based on a large sample of data points, these parameters are discussed in the potential problems section below.

6.3.2.2 Custom SW Estimation

Reifer Consultants, Inc. strongly recommend performing two estimates. To our knowledge, this has not been done.

6.3.2.3 Operational Cost Estimation

With regard to operations costs, a key issue is the level of automation. Often, alternatives being traded differ with respect to the level of automation assumed. To correctly evaluate the trade in such cases, it must be possible to account for the different levels of operations costs in the overall cost estimates. If such costs cannot be estimated in some semi-automated way, that aspect of the trade is generally just ignored. Hence, some degree of automation in computation of operations costs is highly desirable. This is the primary deficiency in the estimation of operations costs; it is highly manual, and, as a result, is only conducted once in each development phase. A separate method which estimates approximate operations costs in a semi-automated fashion is needed to facilitate trade analysis. The partial operations cost estimation included in the COTS cost estimation model does not satisfy this criteria since it is not related to level of automation; it is strictly driven by the number of processors being estimated by the model.

6.3.3 Potential Issues

6.3.3.1 COTS HW and SW Estimation

Within the COTS estimation model, the utilized cost/performance decrease per year seems too low. Archive is particularly low, where the estimated value is 31% per year compared to the 5% being used. The few data points measured definitely produced higher values, which would reduce the cost estimate overall. Further, experimentation with the values suggest that one must use historically inaccurate numbers to get a percent change per year comparable to those used with this model. However, it is necessary to examine a larger database to be sure.

As previously discussed, the generic classes within the COTS Procurement Cost Model may equate to old technology. One sample output of that model contained entries for 780 MB disks and 3480 archive technology. Both of these technologies are cost drivers.

The values of price/performance decrease used in the COTS Procurement Cost Model are unknown at this time. However, it seems likely that the percentages used in the COTS estimation model were derived from EDS data. Therefore, the percentages used in the COTS Procurement Cost Model are assumed to be similar. Of course, it is also unknown which set of curves were utilized within the COTS Procurement Cost Model. If however, similar values were used, the COTS hardware and software cost may be overestimated.

6.3.3.2 Custom SW Estimation

There are two primary issues regarding custom software estimation. First, is the technique valid or not? The analysis is not complete at this time, however, there may not be a sufficient database or sufficient understanding of the parameters that drive software size when using OOD. Software development in C++ does not have a long history to draw upon. Realistically, use of object-oriented languages is only now becoming commonplace. Accordingly, relatively little has been published on the topic. Further, some of the analyses that have been done have derived relationships between function points (or analogies) and SLOC. Very little has been done to directly relate object-oriented design entities to SLOC.

The second issue regarding custom software estimation is whether the multipliers used are correct or not. The analysis so far is mixed. When compared with SAMPEX and FAST, the multipliers seem low, even when the SAMPEX and FAST parameters are increased by a factor of four to account for the level of maturity of any design at SDR. However, when compared with the parameters estimated for the NIHCL, the figures are approximately correct. Given that the NIHCL is an unusual set of software, highly optimized for reuse, it is unlikely that it provides an appropriate standard of comparison. When approached from another point of view, “Is the total estimate is credible or not?”; the answer is again mixed. When compared with analogous systems of the past, the total SLOC estimate is believable, and perhaps even somewhat low. However, in this era there is a much higher availability of COTS tools which either directly supplant code in previous systems or make the development of that code much more efficient. The area of user interface development is a prime example. Today’s development tools provide the means to specify the interface graphically and then generate the code that implements the interface automatically. Clearly, this is an area that requires further study. PDR estimates should represent a major improvement in quality and provide an excellent basis for more detailed analysis.

6.3.3.3 Operational Cost Estimation

It is not clear that operations staffing has fully accounted for the wall clock time to perform media production. A standard multiplier of five minutes handling time for each piece of media generated is being used. Whether this can be met not only depends on the level of automation but the number media generation units being employed, since up to 30 minutes is the typical time to generate a CD-ROM, and one of the steps the operations staff must complete. Within the Data Distribution Facility at GSFC, which has been performing this function for several years, it has been found that the operator involvement with a single piece of media is typically 15-20 minutes. This includes loading the media for recording, unloading and reloading the media in another device for verification, associating the media with its shipping documentation and labels, and packing the media for shipping. Even if the verification step is skipped, the time is still in excess of 10 minutes. Another aspect of this issue is that in order to keep equipment costs to a minimum, it may be necessary to staff media production activities on the second and third shifts, increasing the average rate.

During the interview process, it was stated that the standard multiplier for the cost of a man-year was $100K. It is unknown whether that value was or is being used in the estimation of operational costs. Using representative staffing levels, it is possible to obtain an average cost per man year across all labor categories of between $80K and $95K, depending on the salary levels and indirect rate assumptions. However, since this work is predominated by lower salary levels and on-site labor, the costs should tend toward the lower end. Further analysis is required to determine whether use of the $100K figure represents a problem or not. The labor categories, number of staff in each category, hourly rate assumption for each category, and indirect rate assumptions used in computing the average annual cost per man-year are presented in Appendix D. In any case, to perform this type of cost estimation, the only truly accurate method is to use separate rates for each labor category and separate indirect rates for on-site and off-site personnel.

6.4 Conclusions

The overall conclusions from the analysis performed are:

- There is no life-cycle cost model;
- The trade space is too restricted and insufficient in scope;
- The software size estimation process is unproven; and
- The parameters being used are too conservative.

One result of the inability to easily compute life-cycle cost is that trade studies take a long time to complete, leaving little, if any, chance for examination of alternate trades. Rarely does one find the optimum trade among the first set of alternatives considered. Instead, it is necessary to view the results, come to some conclusions about what aspects warrant further investigation, and analyze the resulting (new) alternatives. The most common problem in system design, which results in excessive costs or broken schedules is that not enough alternatives were considered. Problems are encountered and/or envisioned, but alternative solutions to those problems are not found because they are never considered. Lack of ability to analyze alternatives quickly is a leading cause of this type of behavior.

Another result of the inability to easily compute life-cycle cost is that trade analysis are often performed on the basis of a subset of the cost data, such as COTS hardware and software. This can be very misleading and result in design decisions that actually increase cost rather than reduce it. These behaviors may, in fact, be occurring on this project. Within the "PDR Modeling Plan", there is no mention of computing life-cycle cost within the Cost section. The modeling schedule does not show any indication that a progressive refinement of trades will occur.

Finally, as noted in previous sections of this report, the parameters being used in the various models are probably too conservative. Within COTS hardware and software estimation, the prime example is the percentage decrease in cost per unit of performance per year. Within custom software estimation it is the SLOC multipliers per operation, and within operations cost estimation it is the cost per man-year. Some degree of conservatism is warranted in the early stages of the life-cycle. However, overly conservative estimates can also lead to problems such as unnecessary descoping of systems and selection of less than desirable operations concepts.

6.4.1 Technical Integrity

Technical integrity was evaluated in three areas: traceability to requirements; engineering quality; and testability. Overall, traceability to requirements is mixed. The only requirements which explicitly pertain to cost modeling are found in the Statement of Work for the development contractor. These requirements pertain to life cycle cost estimation, including the various elements of cost examined in this study. In general, there is no overall model which computes life cycle cost. Rather, outputs from a series of models must be manually accumulated. The satisfaction of each requirement in the statement of work is presented in Exhibit D-3.

Engineering quality is good overall. The analytic ability applied is sound and mature. There is a range of difficulty with regard to testability, however, as testability is generally difficult. Our technical integrity evaluation for each modeling area is presented below.

6.4.1.1 COTS HW and SW Estimation

This modeling area would be moderately testable if the models were provided. The models are implemented as a series of spreadsheets, making the inspection of the model structure relatively easy. The primary difficulty in testing these models is that the spreadsheets are not tied together in an automated way, leaving a set of questions unanswerable without further discussion with the model developer/operator. Without the models in hand, assessing the testability is very difficult, therefore requiring development of independent estimates to validate the model outputs.

With respect to engineering quality, the approach underlying the COTS Cost Estimation Model and the COTS Procurement Cost Model is reasonable and sound. In general, good engineering judgment has been applied. Both models depend on the validity of the input data; i.e., the workload data for COTS Cost Estimation Model, and the bill of materials for the COTS Procurement Cost Model. The difficulties have to do with the degree of automation and scope. Degree of automation has been discussed thoroughly above. However, it should be noted that the scope of the COTS Cost Estimation Model is also lacking. Since this model is designed for trade analysis, it should have the ability to estimate a wider range of operations costs and custom software costs at a high level to permit them to be considered in the trade. Based upon these factors, we do not feel that this model can be considered to be fully mature. Therefore, we have assigned a maturity rating of 2, Somewhat Limited Maturity, to this modeling area.

6.4.1.2 Custom SW Estimation

The custom software estimation approach is moderately testable given the models. The approach uses automation which allows individual parameters to be selected and examined. The primary difficulty lies in assessing the counts of operations per object class. Performing that analysis requires a non-trivial degree of familiarity with the software design. Without the models, the testability becomes poor, requiring a count of operations for every class, as well as a complete schedule analysis to arrive at an independent estimate for comparison with the model output.

Regarding engineering quality, the technique for estimating code size is reasonable. However, it is unproven and not known to be calibrated. In comparison to other methods such as function points or object-oriented units (as described in Section 6.3.1.2.2 above), the method seems to be underparameterized. That is to say, it only counts one type of entity (i.e., operations), and uses that count as the basis of the estimate. Object classes and external entities, which probably would have different multipliers, are not explicitly represented. With respect to estimation of effort and schedule given a SLOC count, a typical COCOMO-like approach was used. Since the parameters used there were not made available, nor was the staffing mix or cost per hour, the quality of the final cost estimate cannot be judged at this time.

Regarding maturity, this modeling area is mixed. We have concluded that because we have not seen any evidence that the method for estimating code size has been well studied and is thoroughly calibrated, that it of limited maturity at this time. When additional information is available to us in this area, our assessment will change to reflect the additional insight gained. The method for estimating effort and schedule, on the other hand, is fully mature. However, since the estimation of effort and schedule depend on the size estimate, we have concluded that the entire modeling area must be regarded to be of limited maturity.

6.4.1.3 Operational Cost Estimation

The testability of this modeling area is limited. There are two main reasons for this. First, there are no established models that can be acquired and carefully studied to evaluate the methods being used. Hence, the primary approach must be to validate this modeling area through independent analysis and comparison of results. This leads to the second reason for limited testability; a high degree of knowledge of phasing of automation is required to perform an independent analysis of operations costs.

The overall conclusion regarding the engineering quality of the operational cost estimation area is that the approach is reasonable. The cost estimation process is thorough; all of the elements and major staff functions have been included. There does not appear to be any overlap between this model and the other model areas. There appears to be a conscious effort to keep staffing at minimum levels. The major uncertainty at this time is how cost was estimated for staffing. There is also a concern about the lack of automation of this modeling area, leading to overly subjective cost estimates and an inability to support trade analysis. Based upon these factors, we do not feel that this model can be considered to be fully mature. Therefore, we have assigned a maturity rating of 2, Somewhat Limited Maturity, to this modeling area.

6.4.2 User Satisfaction

6.4.2.1 COTS HW and SW Estimation

The potential users of these models are the engineering staff and the instrument teams (to support their trade-off analysis). The COTS cost estimation model satisfies the basic needs of the engineering staff, but limits the scope of their work due to lack of automation. The model implementation does provide maximum flexibility. Flexibility is important, because of the analyses required and because the ground rules are in a constant state of flux. However, this approach is dependent on the knowledge, analytical skills, and thoroughness of the operator. This has two effects. First, it makes the model essentially unusable by others. Second, it makes the execution of the model very time consuming. Further, the model currently contains embedded rate information, making its release to outside entities problematic. The COTS cost estimation model need not contain this type of information in order to support trade analyses. The accuracy of the estimates produced by the COTS procurement cost model, which is its most important requirement, could not be evaluated in this study due to a lack of information.

6.4.2.2 Custom SW Estimation

The method or model being used for custom software estimation certainly satisfies the needs of the engineering team. However, this model is inaccessible to all groups outside the engineering team due to the degree of detailed knowledge required. A simpler method is needed to support trade analysis and lifecycle cost estimation.

6.4.2.3 Operational Cost Estimation

As with the COTS hardware and software estimation techniques, the methods being used for operational cost estimation meet the basic need of the engineering staff, but limit the types of analyses than can be performed. This model is inaccessible to everyone else due to the degree of detailed knowledge required.

6.4.3 Trends and Projections

There is no indication at this time that the degree of automation of the models or the scope of the models is changing. The specifics of each modeling area are discussed below.

6.4.3.1 COTS HW and SW Estimation

A positive trend in the area of COTS hardware and software estimation is that the estimation is becoming more strongly tied to performance analysis and actual designs. The result will be improved accuracy of the estimates. There is no indication at this time that the parameters used in the cost estimation process are changing or becoming less conservative.

6.4.3.2 Custom SW Estimation

As with COTS hardware and software, custom software estimates generally get better as the design matures. Whereas for SDR it was necessary to estimate numbers of operations, for PDR it will be possible to simply count them based on the object class specifications. Once again, whether the estimates will be valid or not will also depend on the multipliers selected. The current plan is to use the same multipliers for PDR as were used for SDR. This may overestimate the costs since the specificity of the design has improved.

6.4.3.3 Operational Cost Estimation

As with the other modeling areas, as the system design improves, the understanding of the operational roles also improves causing the estimate of operational costs to improve. Serious attention has been given to defining those roles and the phasing of automation during the pre-PDR time frame. The process, however, continues to be very subjective and manual.

6.5 Recommendations

6.5.1 Areas Requiring Further Analysis

6.5.1.1 COTS HW and SW Estimation

As noted above, the methods being used to estimate COTS hardware and software costs are reasonable, in general. The real issue lies in the parameters used within those methods. Examination of percent decrease in price/performance has already raised a question about the parameters being used there. Other parameters, such as the make, model, and assumed costs which make up the starting point for cost for each item, should also be examined. There are three ways to accomplish this. The first is to directly compare the cost parameters being used by HAIS to costs in other databases or those currently available in the marketplace. This requires obtaining detailed cost data from HAIS, which up to this point has been very difficult. The second method is to obtain the bill of materials from HAIS and evaluate the entries to insure current makes and models are being used, and then make an independent estimate of the cost for comparison with the HAIS estimate. Making such an estimate would depend on access to cost data from other sources. One such source is the Small Engineering Workstation Procurement (SEWP) vehicle at GSFC. This vehicle contains equipment from a wide range of vendors (including DEC, SUN, SGI, HP, and others), a wide range of types (computers, peripherals, network equipment, terminals, etc.) and a wide range of performance specifications. An on-line database of model numbers and costs is available. Other contract vehicles exist at GSFC covering archive and media generation equipment. The IV&V team also has working relationships with the major vendors in this area, including R-Squared, E-Systems, Exabyte, Sony, and Phillips, and is familiar with the hardware and software costs. Even if the bill of materials or its associated cost cannot be obtained, a third method is possible. In this case a completely independent estimate would be made based on an independent bill of materials. Such a bill of materials would be generated using performance models developed/obtained for the performance analysis portion of this work. Costs would then be estimated using the independent means described above.

6.5.1.2 Custom SW Estimation

During the course of this study, HAIS did not make the models used to estimate software size and cost available for verification and validation. Hence, that step remains to be completed. As noted above, there may be strong resistance to making those models available. However, without some type of additional analysis it is impossible to determine the degree of accuracy in those modeling results. One method would be to simply obtain the models and scrutinize the parameters used within them. However, based on the generally held opinion that two independent estimates are really needed to arrive at a sound estimate, it would be more desirable to make a second estimate without considering the HAIS models and their parameters. This could be done using one of the alternate methods, such as function points or object-oriented units, to estimate size. Tools may exist to automate this process. If so, such tools would be evaluated and considered for acquisition. Other tools such as Reifer's Softcost-R (which is available to the IV&V team) could then be used to estimate effort, schedule, and cost.

Three factors that can have a dramatic effect on cost are: reuse of existing code; use of COTS packages in place of custom code; and use of development tools to make all custom development as efficient as possible. The design, development strategy, and cost estimation approach should be carefully examined to insure that these factors are being employed to the maximum extent possible.

Finally, the possibility that the Version 0 IMS will become the TRMM Release IMS has surfaced. If this becomes the reality, PDR software size and cost estimates must be examined to insure that this fact has been reflected in those estimates.

6.5.1.3 Operational Cost Estimation

The assumption regarding the number of minutes a media specialist must be involved with a single piece of media, on average, has been identified as a potential problem. This analysis is predicated upon the typical types of designs. However, it is possible that other designs could yield less operations involvement. Hence, alternative system and operations concepts should be developed and analyzed to determine how the overall process might be made more efficient. Rather than relying solely on imagination to arrive at these alternatives, the analysis should first examine how media generation is performed commercially. Several new alternatives should be developed and analyzed based upon the commercial paradigms. While collecting this data, the multipliers used within the commercial context to estimate staffing levels should also be determined.

The numbers of staff and their allocations for the SDR time frame were made available, but were deemed to be invalid. Hence, no time was taken to analyze that data. The number of staff and allocations for the PDR time frame was not available at the time of this writing, and therefore could not be examined. Hence, this step remains to be done.

6.5.2 Solutions to Important Problems

There are two overall problems which can and should be addressed: 1) automation, and 2) availability of cost data. Automation can be addressed by implementing higher level models for doing main trades. As noted before, these models need not produce highly accurate estimates, as long as the estimates are relatively correct. These higher level models could then be connected together to produce a life-cycle cost model. Regarding cost data, it is imperative that this data be validated. There are three possible solutions: 1) the government reviews specific cost data (comparison data could be provided by IV&V contractor), 2) the IV&V contractor signs necessary non-disclosure agreements the obtains and reviews data, and 3) the IV&V contractor makes independent estimates as described above. A discussion of specific problems and suggested solutions within each modeling area are discussed below.

6.5.2.1 COTS HW and SW Estimation

With the exception of the automation issue mentioned above, the problems in this area are all rather simple to solve. The difficulty is in reaching agreement that the solutions should be implemented. In general, the change desired is to use more aggressive parameters in the estimation process. These options already exist in many cases. Where they do not exist, the data needed to compute the appropriate parameter values are readily available. To the extent that old technology entries exist as a basis of cost, these should be updated. Regarding automation, beyond the higher level models discussed above, a greater degree of automation can be achieved by implementing appropriate decision logic and control parameters in the existing spreadsheets. For example, it should be possible to implement flags indicating which of the four estimation techniques is to be used for each cost element, and the source(s) of the required data. A standard format of the input data would also need to be adopted. The details would require some time to work out and implement, but it is nevertheless imminently doable.

6.5.2.2 Custom SW Estimation

The primary problem in this area was the lack of an independent estimate. Hence, such an estimate should be performed.

6.5.2.3 Operational Cost Estimation

The lack of automation in this area can also be addressed by adopting quantitative methods for computing numbers of staff. Even if these numbers are only deemed to be useful as guidelines, this would remove much of the subjectivity and seriously reduce the time required to arrive at an estimate.

6.5.3 Risk Management

The acquisition method being used for EOSDIS gives rise to certain challenges regarding cost estimation. In general, with this type of contract the goal must be to insure that the development contractor is not overestimating the cost. This is somewhat contrary to risk management, in which the cost must be increased to reflect uncertainties in how the system will be built and how long it might take. The solution is to do both. Development contractor estimates should be carefully scrubbed and examined to make sure that all estimation techniques and basic cost data are correct. Uncertainty factors must then be added to account for risk. Potential risks for each area of cost estimation are discussed below.

6.5.3.1 COTS HW and SW Estimation

The main risk in this cost estimation area is that some of the fundamental design concepts may fail, resulting in dramatic changes in the types of hardware required to implement the system. One example is that the distributed data and distributed control may not be fully achievable. In this case, one solution would be to centralize the database server, requiring much more powerful hardware and increased cost. Another example is that the existing communications infrastructure may be insufficient, requiring additional dedicated hardware and/or leased lines and additional cost.

Mitigation of these risks involves doing more prototyping and more modeling, as are already being done. However, risk in the cost estimates can also be mitigated by adding cost based on perceived probability and the cost of the associated changes.

6.5.3.2 Custom SW Estimation

In the custom software estimation area, the main risks are that certain functions may not be available off the shelf, causing an increase in the size of the code. This could also impact schedule and cost by increasing the development time for low-level complex system functions.

Just as with COTS hardware and software estimation, outside of doing the appropriate prototyping, cost risk can be mitigated by adding cost based on the perceived probability and cost of the changes.

6.5.3.3 Operational Cost Estimation

In this cost estimation area the main risks are that certain automated capabilities may not materialize on schedule or in some cases, may not materialize at all, and that multipliers for media handling may not be attainable, resulting in higher than expected operational costs. These risks are mitigated by prototyping and adding cost based on the perceived probability and cost of the changes.

Appendix A: User Model Analysis

This Appendix contains three sections. The first section contains detailed information about the analysis of user demography and research area interests from the user scenarios. The second section contains the listing of the satellite or instrument data requirements from the user scenarios. The third section contains the mapping of the functions or services from the user scenarios to the Level 3 requirements.

A.1 User Scenario Data Analysis Tables

This section lists the numbers of users according to the epoch numbers that were used by HAIS in their user scenario spreadsheet: epoch 1, epoch 2, epoch 3, and epoch 4. These map to the epochs identified in Exhibit 2-5 in the following way:

Appendix A.1	Exhibit 2-5
Epoch 1 (early 97)	Epoch A (Dec 96–Jun 97)
Epoch 2 (early 98)	Epoch C (Jan 98–Jun 98)
Epoch 3 (early 99)	Epoch E (Jan 99–Jun 99)
Epoch 4 (mid-99)	Epoch F (Jul 99–)

Scenario	Investigator(s)	# Steps	Discipline	USGCRP		Epoch 1	(Early 97)	Epoch 2	(Early 98)	Epoch 3	(Early 99)	Epoch 4	(Mid-99)
						min	max	min	max	min	max	min	max
1	Flittner	9	——	——		1	1	14	16	14	22	15	23
2	Holle	17	atm	chs		3	4	70	78	71	111	73	116
3	Strebel/Sellers	6	land	eco		1	1	14	16	14	22	15	23
4	Friend	9	land	eco		1	2	22	24	22	34	23	36
5	Dozier	12	——	——		4	6	105	117	106	165	109	174
6	Garegnani	7	land	eco		26	44	860	960	870	1360	900	1430
7	Baldwin	12	land	eco		13	22	430	480	435	680	450	715
8	Justice	14	land	eco		19	31	602	672	609	952	630	1001
9	Poston	11	land	eco		9	15	285	318	288	451	298	474
10a	Engman	10	land	chs		20	33	645	720	653	1020	675	1073
10b	Wielicki	4	atm	chs		20	33	645	720	653	1020	675	1073
11a	Heinrichs	11	cryo	chs		5	8	144	160	145	227	150	239
11b	Walsh/Khalsa/Kaminski	9	cryo	chs		5	8	144	160	145	227	150	239
11c	Key	10	cryo	chs		5	8	144	160	145	227	150	239
12	Rood/Stobie	9	atm	chs		19	31	602	672	609	952	630	1001
13	Barkstrom	34	atm	chs		8	13	236	264	239	373	247	392
14	Rosenthal	27	land	chs		19	31	602	672	609	952	630	1001
15	Lopez	35	atm	chs		11	18	344	384	348	544	360	572
16	Walstad	10	ocean	chs		28	48	925	1032	936	1462	968	1538
18	Hannawald	5	land	chs		16	26	500	558	506	791	523	831
19	Goyet	8	ocean	bio		5	8	146	163	148	230	153	242
20	Barron	7	land	chs		13	22	430	480	435	680	450	715
22a	Flynn	11	land	solid		1	1	11	12	11	17	11	18
22b	Isacks	8	land	solid		1	1	11	12	11	17	11	18
23a	Lait/Schoberl	5	atm	chs		4	7	125	140	127	198	131	208
23b	Goodman	9	atm	chs		4	7	125	140	127	198	131	208
24	Abbott	12	ocean	chs		11	18	344	384	348	544	360	572
					Total	**272**	**447**	**8525**	**9514**	**8624**	**13476**	**8918**	**14171**

EXHIBIT A-1 Scenario Analysis--Research Areas And Number Of Users

EXHIBIT A-1: Scenario Analysis—Research Areas And Number Of User

Science	Disciplines	
	Number	%
atm	7	26%
cryo	3	11%
land	12	44%
ocean	3	11%
(no pref)	2	7%
total	**27**	**100%**

USGCRP	Research	Area
	Number	%
chs	16	59%
biodyn	1	4%
eco	6	22%
hist	0	0%
humint	0	0%
solid	2	7%
solar	0	0%
(no pref)	2	7%
total	**27**	**100%**

EOS-Funded	Investigators	
	Number	%
chs	396	37%
biodyn	175	16%
eco	149	14%
hist	109	10%
humint	65	6%
solid	75	7%
solar	98	9%
(no pref)	0	0%

Number of Users by Discipline

Discipline	Epoch 1	(Early 97)	Epoch 2	(Early 98)	Epoch 3	(Early 99)	Epoch 4	(Mid-99)
	min	max	min	max	min	max	min	max
atm	69	113	2147	2398	2174	3396	2247	3570
cryo	15	24	432	480	435	681	450	717
land	139	229	4412	4924	4463	6976	4616	7335
ocean	44	74	1415	1579	1432	2236	1481	2352
(no pref)	5	7	119	133	120	187	124	197
total	**272**	**447**	**8525**	**9514**	**8624**	**13476**	**8918**	**14171**

Percentage of Users by Discipline

Discipline	Epoch 1	(Early 97)	Epoch 2	(Early 98)	Epoch 3	(Early 99)	Epoch 4	(Mid-99)
	min	max	min	max	min	max	min	max
atm	25.37%	25.28%	25.18%	25.20%	25.21%	25.20%	25.20%	25.19%
cryo	5.51%	5.37%	5.07%	5.05%	5.04%	5.05%	5.05%	5.06%
land	51.10%	51.23%	51.75%	51.76%	51.75%	51.77%	51.76%	51.76%
ocean	16.18%	16.55%	16.60%	16.60%	16.60%	16.59%	16.61%	16.60%
(no pref)	1.84%	1.57%	1.40%	1.40%	1.39%	1.39%	1.39%	1.39%
total	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Number of Users by USGCRP Res. Area

Res. Area	Epoch 1	(Early 97)	Epoch 2	(Early 98)	Epoch 3	(Early 99)	Epoch 4	(Mid-99)
	min	max	min	max	min	max	min	max
chs	191	315	6025	6724	6096	9526	6303	10017
biodyn	5	8	146	163	148	230	153	242
eco	69	115	2213	2470	2238	3499	2316	3679
hist	0	0	0	0	0	0	0	0
humint	0	0	0	0	0	0	0	0
solid	2	2	22	24	22	34	22	36
solar	0	0	0	0	0	0	0	0
(no pref)	5	7	119	133	120	187	124	197
total	**272**	**447**	**8525**	**9514**	**8624**	**13476**	**8918**	**14171**

Percentage of Users by USGCRP Res. Area

Res. Area	Epoch 1	(Early 97)	Epoch 2	(Early 98)	Epoch 3	(Early 99)	Epoch 4	(Mid-99)
	min	max	min	max	min	max	min	max

EXHIBIT A-2: Analysis Of Scenario Users

Appendix A.2 Satellite/Instrument Data Requirements From The User Scenarios

		TRMM	LAND SAT				AM-1				PM-1									CHE M-1						TOPEX	ADEOS	SWIFS	UARS	RADAR
NO	AVHRR			CERES	MISR	MODIS	ASTER	MOPITT	AIRS	AMSU	CERES	MODIS	MHS	MIMR	ACRIM	SAGE	LIS	SAR	MLS	TES	HIRDLS	GLAS	TOMS	SOLST ICE	SSMI					ALT
1																							X	X						
2																	X													
3	X					X																								
4						X		X						X																
5																														
6			X			X	X																							
7	X		X				X																							
8						X	X																							
9	X																													
10A				X	X	X			X		X							X												
10B				X		X	X																							
11A																		X							X					
11B																									X					
11C	X			X																										
12				X	X	X	X	X																						
13				X																										
14			X																											
15																	X													
16																										X				
18						X	X																							
19						X																						X		
20	X			X		X	X																							
22A					X	X	X																							
22B			X		X	X	X							X								X								
23A	X																								X				X	
23B		X																												
24						X																			X			X		

EXHIBIT A-3: Functions Or Services From User Scenarios Mapped To The ECS Level 3 Requirements

Appendix A3: Mapping Of The Functions Or Services To Level 3 Requirements

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
1	Simple Search	A search based on parameters entered by the user. Multiple data sets and multiple DAACs may be located. Documents may also be located.	IMS-0510 The IMS shall provide tools for research planning and data search, to include at a minimum: a. Data acquisition schedules and plans b. The capability to map specified geophysical parameters to the appropriate instrument and/or Standard Product c. Descriptive information on instruments and geophysical parameters available in Standard Products d. Climatology information e. Phenomenology information f. Geographic reference aids g. Spacecraft location projections IMS-0530 The IMS shall provide document text search. IMS-0575 The IMS shall provide the capability to search across multiple data sets for coincident occurrences of data in space and/or time and any other attribute(s) of metadata.	All scenarios excepting Scenario 20 (8907/16209)	Search for groups of related data products and/or parameters for base mapping purposes (#10A) URDB # 626 will be functionally supported by ECS.(limited to large scale features). Simultaneous search requested for #10A for QA statistics is categorized as implementation detail	# URDB 626 refers to search for a number of related products at the same time. Status of this requirement is classified as 'screening'.
2	Match-up Search	The user already has retrieved data or metadata and then asks for additional data or metadata that matches the previous set. For example, a user first locates a list of data sets which contain his parameters. He next asks for Guide information about those data sets. Or, a user is examining a product and next asks to view the product history.	SDPS-0095 generally addresses this requirement. SDPS0095 The SDPS shall provide science user interfaces that are individually tailorable including settable preferences, user defined keywords, query save capabilities, and screen layout preferences.	Scenario Nos. 2, 3, 4, 6, 7, 8, 9, 10A, 11A, 11B, 11C, 12, 15, 19, 20, 23B (5737/10472)	HAIS have indicated that this requirement is being implemented, subject to the hardware requirements within the limits for this functionality ,-the necessary storage may be required at the user facility. URDB #610 and # 538 which are relevant are classified as design issues.	This functionality is required by a large number of scenarios and it should be clearly specified in F&PRS
3	Coincident Search	The user wants to locate two or more data sets simultaneously that match his criteria.	IMS-0575 The IMS shall provide the capability to search across multiple data sets for coincident occurrences of data in space and/or time and any other attribute(s) of metadata. IMS-1780 The IMS shall respond to each user session operation within the time period specified in Table 7-4 with the specified rate of IMS operations.	# 4 and #20 (525/945)	——	The requirement to search multiple inventories simultaneously is included in Table 7-4 referred to in IMS-1780

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
4	Simultaneous Multiple DAAC Search	A search is issued that requires the system to search at more than one DAAC or other data center. This takes place either because the user query is very general (i.e., "find me all products containing radiative flux") or because products from one instrument may be located at more than one DAAC.	IMS-0575, IMS-0140, IMS-1780 refer to coincident searches, simultaneous sessions and multiple DAAC inventory searches. (IMS-0575 & 1780 see above) IMS-0140 The IMS shall provide the capability for multiple simultaneous sessions – for example, the capability to transition back and forth smoothly between directory search, inventory search, and data visualization. For example, when viewing a directory entry, the user shall have easy access to the corresponding guide (documentation/reference material) and inventory information.	Scenarios 3, 4, 6, 8, 9, 10A, 11A, 11B, 11C, 12, 16, 18, 19, 20, 23B (6493/11687)	——	No issue identified
5	Save Query Results to file for later use	A user wants granule IDs or other metadata from multiple searches saved to a file so that when he is ready to order, he can retrieve the file and select from it.		User scenarios 1, 4, 7, 9, 10A, 14, 23B. (2451/ 4441)	——	SDPS-0095 provides for query saving but no requirement exists for saving the metadata from query results. URDB #400 classified as design consideration, addresses this requirement
6	Spatial Subsetting	The user does not want data covering the entire area of the globe, but desires data covering some smaller portion of the globe.	DADS1475 Each DADS shall provide tools to the users to perform: a. Format conversion of EOS data b. Subsetting c. Compression (lossy, lossless) d. Data transformation e. Subsampling	User scenarios 2, 3, 6, 7, 10A, 10B, 11A, 11B, 11C, 12, 13, 14, 15, 16, 18, 19, 20, 22A, 22B, 23A, 23B, 24. (8630/15681)	The capabilities provided may be limited by the construction and format of the data product and not by ECS capabilities.	No issue identified

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
7	Temporal Subsetting	The user does not want all of the data pertaining to his area of interest, but only data within a specified time span.	DADS-1475 and IMS-0720 partially meets the requirement DADS-1475 See above IMS-0720 The IMS shall provide the capability to request data products which are processed ad hoc in response to user requests for subsetting, subsampling, or averaging within a granule based on defined criteria to include: a. Geographical location (x, y, z - spatial with rectangular boundaries) b. Spectral band c. Time	User scenarios .2, 6, 7, 8, 10A, 10B, 11A, 11B, 12, 13, 14, 15, 16, 18, 19, 20, 22A, 22B, 23B (8602/15631)	The capabilities provided may be limited by the construction and format of the data product and not by ECS capabilities.	No issue identified
8	Parametric Subsetting	The user does not want all of the parameters that make up a product.	DADS -1475 and IMS-0720 DADS-1475 See above IMS-0720 The IMS shall provide the capability to request data products which are processed ad hoc in response to user requests for subsetting, subsampling, or averaging within a granule based on defined criteria to include: a. Geographical location (x, y, z spatial with rectangular boundaries) b. Spectral band c. Time	User scenarios 2, 4, 6, 8, 10A, 10B, 11C, 12, 13, 15, 16, 18, 19, 20, 22A, 22B, 24. (7805/14197)	The capabilities provided may be limited by the construction and format of the data product and not by ECS capabilities	No issue identified
9	Spectral Subsetting	The user does not want all of the spectral bands contained in a product.	DADS -1475 and IMS-0720 (See above)	User scenarios 1, 14, 22A (712/1311)	The capabilities provided may be limited by the construction and format of the data product and not by ECS capabilities	No issue identified
10	Subset QA statistics	The user does not want all of the QA statistics; he wants the statistics covering his time span and area of interest only.	DADS -1475 and IMS-0720 (See above)	User scenario 8 (700/1260)	HAIS considers this as an implementation detail of the current SDS client support services.	Subsettiing is required to be provided for data and meta data, where QA statistics are expected to be available.

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
11	Save Subsetted data for later "bulk" retrieval	The user subsets his data in space, time, etc. and now asks that the data be stored on ECS as a group to be retrieved at a later date.	IMS-0480 The IMS shall allow the user to store documents in the ECS.	User scenario 10B (750/1350)	URDB # 615 addresses this requirement. this has been accepted.	Requirement should be included in F&PRS (Provision however exists for the user storing documents in ECS ,IMS-480)
12	Save list of lat./long coordinates....	The user subsets a large image into several smaller ones. He would then like to return to the large image to process the data in his areas without having to redraw them.	——	User scenario 14 (700/1260)	——	No requirement in F&PRS
13	Access info on non-EOS data	The user wants information (not the data itself) regarding a data product that did not originate from an EOS instrument.	SDPS-0020 The SDPS shall receive EOS science, engineering, ancillary, and quick-look data from the EDOS, the SDPF, and the IPs, and non-EOS data, in situ data, associated algorithms, documentation, correlative data, and ancillary data (as listed in Appendix C) from ADCs, EPDSs, and ODCs DADS0145 Each DADS shall be capable of receiving from the ADCs, at a minimum, the following for the purpose of product generation: a. L0-L4 equivalent data sets b. Metadata c. Ancillary data d. Calibration data e. Correlative data f. Documents g. Algorithms DADS-0260 Each DADS shall receive non-EOS correlative and ancillary digital data.	User scenarios 1, 3, 4, 10A (791/1453)	——	No issue identified

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
14	Access info on EOS and non-EOS algorithms	The user wants information (not the algorithms themselves) about algorithms used to produce either an EOS product or a non-EOS product.	SDPS 0020 (See above) SDPS0090 The SDPS shall interface with the PIs and the other science users to support the development and testing of data product algorithms and QA of produced data products. DADS 0140 Each DADS shall receive from other DAACs, at a minimum, the following for the purpose of product generation: a. L0-L4 b. Metadata c. Ancillary data d. Calibration data e. Correlative data f. Documents g. Algorithms	User scenarios 1, 4, 8, 9, 11B, 11C, 13, 15 (2464/4464)	—	Algorithms are available at DAACs (DADS-0140) for supporting product generation. We presume they are available to all users.
15	Access Electronic Journal	The user wants to access an electronic journal stored on ECS.		User scenarios 1, 5 (121/247)	—	IMS-480 is for providing storage of “documents” in the ECS, these are “documents” and reference materials about data sets. No requirement exists in F&PRS
16	Integrated Browse (text)	The user browses textual information interactively (on-line). This information is in the form of documents, guide information, or algorithm descriptions, or information regarding non-EOS data sets.	IMS-0530 The IMS shall provide document text search.	User scenarios 1 through 9 10B, 11B, 11C, 13, 15, 20, 24.	—	No issue identified

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
17	Cut parts of documents and save to file	The user wants to cut parts of documents and have all of the pieces saved to one file which he can then receive all at once.	——	User scenario 13 (274/494)	——	No requirement exists in F&PRS
18	FTP Browse	The user wants browse data products delivered non-interactively.	IMS-1510 The IMS data visualization toolkit capabilities shall be portable and execute on ECS supported workstations and appropriate ECS facility computers. IMS-1520 The IMS toolkit software shall provide data visualization tools to assist the investigators to perform the following functions, at a minimum: a. QA/Validation of products generated by the PGS b. Algorithm development c. Calibration functions, parameter verification, and anomaly detection d. View subsetted, subsampled, and summarized data whenever associated inventory information is displayed.	User scenario 11C, 23B (312/561)	——	No issue identified
19	Send Browse Products on medium	The user wants his browse products delivered on a medium.	IMS-0590 The IMS shall provide the capability to distribute information: a. On-line (i.e., over a network) b. Off-line (hardcopy or off-line data media)	User scenario 23B (145/261)	——	No issue identified
20	Integrated Browse (data)	The user wants to browse data while on-line.	SDPS0092 The SDPS shall support science user development of new search techniques that dynamically browse the data and metadata. IMS-0625 The IMS shall provide bi-directional interoperability between ECS and V0 for access to the inventory metadata, guide information, and browse products via level III catalog interoperability as specified in ICDs.	User scenarios 7, 8, 10A, 10B, 11A, 11C, 12, 13, 14, 15, 16, 19, 22A, 22B (6376/11478)	——	No issue identified
21	Integrated Browse (non-EOS data)	The user wants to browse non-EOS data while on-line.	SDPS0092 and IMS-0625 (Please see above)	User scenarios 7, 14, 15 (1600/2880)	——	No issue identified
22	Display multiple data sets simultaneously	The user wants to view more than one data set at a time.	——	User scenario 15 (400/720)	——	No requirement exists in F&PRS

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
23	Animation	The user wants to view several "frames" of data quickly, so that it appears as a "movie loop".	IMS-1530 The IMS data visualization toolkit shall provide the capability to visualize data in raster and vector formats and to visualize animated products.	User scenario 14, 15 (1100/1980)	Animation/movie-loop capabilities are expected to be implemented in later systems and not in the TRMM Release. The screening team (URDB) proposes that any ECS requirement for visualization should extend to only browse etc. Utilities for in-depth should be provided by the user with his own resources	Subject of Trade/study
24	Display "product coverage" map	The user wants to view a map that depicts areas which are covered by his data products of interest.	——	User scenario 18 (581/1046)	——	No requirement exists in F&PRS
25	Ingest user software or file	The user wants the system to accept either a program the user has written or a file containing information to be used when he is accessing the system.	——	User scenario 14, 16, 19, 22A, 22B, 23A, 23B (2258/4066)	The request is an implementation detail. However the current thrust of the Client Subsystem design in the SDS allow for use of COTS software, and ECS will provide an extensive library of APIs for interaction with ECS resident data.--Trade studies on user supplied data are currently in progress.	URDB #639 design consideration Trade studies under progress No requirement exists in F&PRS.
26	Create and Display 3-D plot	The user specifies data and wishes to view it as a three dimensional plot on-line.	IMS-1540 The IMS toolkit software shall provide the capability to generate, at a minimum: a. Two-dimensional plots (x-y plots, scatter plots, profiles, histograms) b. Three-dimensional plots c. Contour plots d. Three-dimensional surface diagrams	User scenario 16 (1075/1935)	——	No issue identified

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
27	Create and Display X-Y plot	The user specifies data to be used to create an X-Y plot so that he can view it on-line.	IMS-1540 (see above)	User scenarios 13, 22A (286/516)	——	No issue identified
28	Create and Display new images	The user specifies data to be used to create a new image so he may view it on-line.	IMS-1550 The IMS toolkit data visualization tools shall provide capabilities for image manipulation (e.g., pan, zoom, color, contrast).	User scenarios 13, 14 (974/1754)	——	Visualization capabilities under study.
29	Create and Display contour plot	The user specifies data to be used to create a contour plot that he may view on-line.	IMS-1540 (see above)	User scenario 15 (400/720)	——	No issue identified
30	Create and Display scatter plot	The user specifies data to be used to create a scatter plot that he may view on-line.	IMS-1540 (see above)	User scenario 20 (500/900)	——	No issue identified
31	Manage/Save data created by a user process	The user wants ECS to save data created by any of his processes, whether the data was created with human interaction or not, and manage the data so he may locate it in the future.	——	User scenarios 14, 16, 18, 19, 22A (2537/4568)	URDB #615. The capability may be an implementation detail due to both cost and security issues but is within the scope of the ECS project. Presuming that the amount of storage required is not beyond the available resource envelope, ECS should support the function. Another concern in this request relates to the high data transfer rate.	URDB #615, closed-(Existing requirement) To be included in F&PRS

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
32	Ingest/Archive user results created by processes external to ECS	The user wants ECS to archive results of his research for use by others.	DADS0450 Each DADS shall provide storage, at a minimum, for the following scientist provided data: a. Special data products b. Associated correlative data sets c. Associated ancillary data sets d. Associated calibration data sets e. Research results (articles, algorithms, data sets, software) f. Instrument characterization data sets g. Associated Metadata SDPS0130 The SDPS shall provide the capability for DAACs to exchange data products, browse data, metadata, data quality information, research results, and documentation.	User scenarios 16, 18, 20 (2156/3881)	URDB # 637 Screening. This is regarding archival and storage of non standard products. Although the recommendation is a design issue, ECS Architecture team is paving the way for SCF located products to be advertised and accessible. The recommendation is being analyzed and examined by developers to determine if it is within the resource envelope.	URDB 637 is under screening
33	Trigger process	The user wants a process to take place automatically when he logs onto the system.	——	User scenario 15 (400/720)	URDB #611 design consideration	No requirement exists in F&PRS
34	Access Level 0 data	The user wishes that Level 0 data be accessible by himself and/or his process (software).	——	User scenario 22A (12/22)	URDB # 643 Screening, URDB # 644 Assessment. The request remains as a design issue in order to evaluate those applications' needs against the APIs which are feasible to provide. Policy is needed and is being addressed in the trade study on user supplied processing.	No requirement exists in F&PRS

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
35	Access orbital Model output	The user wishes to know when the instruments will taking future data over his area of interest.	——	User scenario 22B (12/22)	URDB #629 and URDB #630 Screening. ECS plans to provide graphical displays of orbital paths most likely with ground tracks included. Limited GIS capability for visualization may be provided. Access to an orbital prediction model will be provided to authorized users for visualization and planning purposes.	No requirement exists in F&PRS
36	Access data dependency info	The user wishes to retrieve a list of data products that were used as inputs to the products he is interested in.	IMS-0330 The metadata maintained by the IMS shall provide a cross reference that relates science data to the following at a minimum: a. Calibration data, navigation data, and instrument engineering data b. Processing algorithms used for data generation at the PGS c. Software used for data generation at the PGS d. Parameters used for data generation at the PGS e. Input data used for data generation at the PGS f. Data recipients g. The PGS at which the data was processed h. QA and validation data, reports, and algorithms	User scenario 15 (400/720)	——	No issue identified
37	Automatic Notification	The user wishes to receive a message at his facility when certain conditions exist.	——	User scenarios 22A, 22B (24/44)	——	URDB #602 (c) Closed (existing requirement) URDB # 645 Assessment No requirement exists in F&PRS

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
38	Statistical Analysis	The user wishes to perform statistical analysis on data while on-line.	IMS-1570 The IMS toolkit software shall provide statistical analysis capabilities. PGS-1020 The PGS shall provide mathematical libraries including: a. Linear algebra and analysis (e.g., LINPAC, IMSL) b. Statistical calculations (e.g., SAS, SPSS)	User scenarios 9,13, 14 (1305/2350)	URDB # 634 Design issue (User defined statistical packages) Though ECS will not provide sophisticated statistical packages on-line, will not limit the use of user defined statistical packages.	No requirement exists in F&PRS for user defined statistical packages.
39	Regrid data	The user would like the data to be recalculated to conform to a grid cell size larger than the current cell size.	—	User scenario 20 (500/900)	—	No requirement exists in F&PRS
40	Coordinate transformation	The user would like to view or receive the data in a coordinate system other than the one the data currently exist in.	—	User scenario 13 (274/494)	URDB # 632 Screening. Although a number of standard projections will be provided by ECS, final decision regarding the specific projections have not yet been made.	Specific requirement is not available in F&PRS
41	Compute difference between two parameters	The user specifies two parameters and would like the system to compute the difference between them.	—	User scenarios 13, 14 (974/1754)	URDB # 618 Assessment	No requirement exists in F&PRS
42	Compute ratio of two parameters	The user specifies two parameters and would like the system to compute the ratio of one to the other.	—	User scenario 14 (700/1260)	URDB # 618 Assessment	No requirement exists in F&PRS
43	Interactive download	The user wants the data he specifies to be sent to his machine and stored on his local disk while on-line.	—	User scenarios 5, 14 (133/240)	URDB # 622 Design consideration The request remains as a design issue in order to evaluate those applications' needs against the APIs which are feasible to provide. Policy is needed and is being addressed in the trade study on user supplied processing.	Study on user supplied processing. No requirement exists in F&PRS

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
44	Point Instrument	The user would like an instrument to be pointed at his area of interest at a time he specifies.	IMS-1070 The IMS shall provide the capability for users to construct DARS for collection of EOS data which shall contain the following information at a minimum: a. Observation number b. Experimenter identification c. Experimenter address d. Investigation identification e. Scientific discipline f. Observation repetition period g. Tolerance in observation time h. User priority i. Scheduling priority and target of opportunity flag j. Descriptive text k. Location data expressed in terms of longitude and latitude as earliest start coordinates and latest stop coordinates l. Earliest start time m. Latest stop time n. Minimum coverage required o. Maximum coverage desired p. Number of instruments involved in the investigation q. Which instruments are involved in the investigation	User scenarios 10B, 22A, 22B (774/1394)	——	IMS-1070 concerns any Data Acquisition Requests
45	Video Teleconferencing	A user wishes to confer with his colleagues while they are all on-line. All should see the same information on their monitors at the same time.	——	User scenario 10A (750/1350)	URDB # 525 Design issue. Because of the uncertainty of communications capabilities and cost, this record is statussed as a design issue and scheduled for re examination in 1995. It would seem that ECS would probably be required to support this from the communications protocol aspect only.	No requirement exists in F&PRS

NO	Functions/Services	from user scenarios	Applicable ECS Level 3 requirements	User scenarios and users (min/max)	Status (HAIS)	IV&V Comments
46	Compute order cost	The user would like to know the cost of the order before final order submission.	IMS-1340 The IMS shall, using information provided by the SMC, provide the capability for users to preview billing costs for EOSDIS data products prior to order submission. IMS-1350 The IMS shall provide the capability for users to preview billing costs, which are based upon MOUs with the ADC and non-EOSDIS data centers, prior to ADC and non-EOSDIS data product order submission. DADS0525 Each DADS shall accept updates/cancellations of data order requests.	User scenarios 1, 2, 3 (16/204)	URDB # 627 Design consideration.	URDB 627 addresses some details of the requirement.
47	Order from a saved results list or file:	The user would like his saved list to be displayed so that he may select products from it while he is ordering. Or, the user would like the system to look in a file where he has predefined the products of interest, the geographic area, the temporal coverage and other required ordering information.	IMS-0050 The IMS shall provide the capability for users to define and modify user profile information, to include at a minimum: a. User electronic address b. Data distribution media c. Data distribution address d. User expertise level e. Default query parameters f. Terminal characteristics g. Technical specialty	User scenarios 1, 4, 7, 22A, 22B (549/1018)	URDB # 611 design consideration This functionality requires that the user profile data file is editable by the user. ECS provides for this capability as stated in requirement IMS-0050.	Details of the requirement are in URDB 611 (Design consideration)
48	Standing Order	The user would like to receive regular shipments of data by submitting only one order. The shipments are sent automatically (does not require action on his part) either at regular intervals or as available.	F&PRS provides for standing orders and is addressed at a number of places. For example: IMS-0740 The IMS shall provide the capability for users to generate and update requests for one-time orders or standing orders for the DADS to distribute DADS archive holdings to include, at a minimum, Standard Products, Standard Product software, EOC historical data, spacecraft housekeeping and ancillary data, and engineering data.	User scenarios 6, 7, 8, 10A, 10B 15, 20, 22A, 23A, 24 (5157/9283)	The specific requirements of the scenarios are in URDB #609 and # 613 which are classified as design issues	Specific requirements are under design consideration.
49	Order results of a user process run on ECS	The user orders data that has been created by his uploaded software using ECS processing power	——	User scenarios 16, 19 (1244/2240)	URDB # 639 design consideration.	No requirement in F&PRS

EXHIBIT A-4: Mapping Of Functions Or Services In The User Scenarios To ECS Level 3 Requirements

Appendix B: PRODUCTION MODEL ANALYSIS DETAIL

B.1 AHWGP Product Availability Scenarios

The Science Operations Concept by Wharton and Myers (discussed in Section 4.1) described a way of baselining, supporting, configuring and delivering science data products through a process which allows the scientists to prioritize the system and allocate a baselined capability to products most in demand, most ready and best performing in the system. In support of this operations concept, the AHWGP created Product Availability Scenarios for TRMM and AM-1 instruments. These scenarios are shown in Exhibit B-1. These scenarios describe anticipated demand for each product processing for each quarter of a year starting from 1st quarter of 1997 to the 4th quarter of 2002. It contains information provided by the Instrument Teams on product maturity, external product needs, spatial coverage and time coverage of standard products to be generated with each instrument data The AHWGP also provided phasing of processing and volume loads for each product at various epochs (in quarterly time intervals from launch to year 2001) to assist in the development of the ECS Production Model. The HAIS PDR Technical Baseline Attachment C in EDHS Community Access Internet Server contains that information in the following documents (Excel spreadsheets) which describe these phasing scenarios: Processing Timelines, Volume Timelines, File Descriptions, and Processing Descriptions (see Appendix F for specific versions used).

We have investigated the adequacy of the Production Model inputs in terms of availability of various instrument data at different epochs and analyzed consistency of data in the Excel spreadsheets with those in AHWGP Product Availability Scenarios and the ATBD requirements. Since the data definition for MOPITT is at a more mature state than most other instruments, we have chosen to focus on MOPITT data as a representative precursor to subsequent analyses. The analysis for MOPITT is presented in Section B.2 The Product Availability Scenarios table contains sequence of letters with following meanings:

1st field:	Product Maturity	N = standard product not available V = undergoing validation, users beware A = available for general use, science team certified
2nd field:	Extent of Parameter Generation within Product	P = partial F = full
3rd field:	Spatial Coverage	R = regional, such as for initial algorithm proving I = intermediate, moderately regular coverage G = global
4th field:	Temporal Coverage	S = sporadic: only a few, irregular times in a month I = intermittent: regular, moderately frequent sampling C = continuous: large fraction of possible samples taken U = user determined as explained in notes below
All fields:		X = designation not applicable for that field.

The time sequence interval is by quarter of a year.

ASTER

Product ID	Platform	Prod. Level	3rd-97	4th-97	1st-98	2nd-98	3rd-98	4th-98	1st-99	2nd-99	3rd-99	4th-99	1st-00	2nd-00	3rd-00	4th-00	1st-01
AST01	AM1	1A					VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST02	AM1	1B					VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST03	AM1	1B					VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST04	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST05	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST06	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST07	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST08	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST09	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST10	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST11	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST12	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST13	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU
AST14	AM1	2					VFUU	VFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU	AFUU

CERES

Product ID	Platform	Prod. Level	3rd-97	4th-97	1st-98	2nd-98	3rd-98	4th-98	1st-99	2nd-99	3rd-99	4th-99	1st-00	2nd-00	3rd-00	4th-00	1st-01
Common Resources																	
CERX06 [MOA]		3	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
[ERBE ADMs] ERBE		3	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
[TRMM ADMs] TRMM		3	NFGC	NFGC	NFGC	NFGC	NFGC	NFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
[TRMM and TRMM &		3			NFGC	NFGC	NFGC	NFGC	NFGC	NFGC	NFGC	NFGC	NFGC	NFGC	AFGC	AFGC	AFGC
AM1 ADMs] AM-1																	
TRMM-Only																	
ERBE-Like Products																	
CER01 [BDS]	TRMM	1B	VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
CER02 [ES-8]	TRMM	2	VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
CER03 [ES-9]	TRMM	3	VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
CER13 [ES-4]	TRMM	3	VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
CER14 [ES-4G]	TRMM	3	VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
Radiative Flux and Cloud Products																	
CER11 [SSF]	TRMM	2	VPGS	VPGS	VPGI	VPGI	VPGI	VPGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC
CER16 [CRH-V]	TRMM	2	VPGS	VPGS	VPGI	VPGI	VPGI	VPGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
CER04 [CRS]	TRMM	2	VPGS	VPGS	VPGI	VPGI	VPGI	VPGC	VPGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC
CER05 [FSW]	TRMM	3	VPGS	VPGS	VPGI	VPGI	VPGI	VPGC	VPGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC
CER07 [SYN]	TRMM	3	VPGS	VPGS	VPGI	VPGI	VPGI	VPGC	VPGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC
CER08 [AVG]	TRMM	3	VPGS	VPGS	VPGI	VPGI	VPGI	VPGC	VPGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC
CER15 [ZAVG]	TRMM	3	VPGS	VPGS	VPGI	VPGI	VPGI	VPGC	VPGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC	APGC
TOA and Surface Radiation Flux Products																	
CER12 [SFC]	TRMM	3	VPGS	VPGS	VPGI	VPGI	VPGI	VPGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
CER06 [SRBAVG]	TRMM	3	VPGS	VPGS	VPGI	VPGI	VPGI	VPGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
AM1-Only																	
ERBE-Like Products																	
CER01 [BDS]	AM1	1B				VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
CER02 [ES-8]	AM1	2				VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
CER03 [ES-9]	AM1	3				VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
\CER13 [ES-4]	AM1	3				VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC

CERES (Contd)

Product	Platform	Prod.

ID		Level	3rd-97	4th-97	1st-98	2nd-98	3rd-98	4th-98	1st-99	2nd-99	3rd-99	4th-99	1st-00	2nd-00	3rd-00	4th-00	1st-01
CER14 [ES-4G]	AM1	3				VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
Radiative Flux and Cloud Products																	
CER11 [SSF] AM1	AM1	2				VFGS	VFGI	VFGI	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
CER16 [CRH-M1]	AM1	2				VFGC	VFGC	VFGC	VFGC	VFGC	VFGC	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
CER04 [CRS]	AM1	2				VPGS	VPGI	VPGI	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
CER05 [FSW]	AM1	3				VPGS	VPGI	VPGI	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
CER07 [SYN]	AM1	3				VFGS	VPGI	VPGI	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
CER08 [AVG]	AM1	3				VFGS	VPGI	VPGI	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
CER15 [ZAVG]	AM1	3				VFGS	VPGI	VPGI	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
TOA and Surface Radiation Flux Products																	
CER12 [SFC]	AM1	3				VFGS	VPGI	VPGI	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
CER06 [SRBAVG]	AM1	3				VFGS	VPGI	VPGI	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
TRMM & AM1 Combined																	
ERBE-Like Products																	
CER03 [ES-9]	T&A1	3				AFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
CER13 [ES-4]	T&A1	3				AFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
CER14 [ES-4G]	T&A1	3				AFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
Radiative Flux and Cloud Products																	
CER07 [SYN]	T&A1	3				VPGI	VPGC	VPGC	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
CER08 [AVG]	T&A1	3				VPGI	VPGC	VPGC	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
CER15 [ZAVG]	T&A1	3				VPGI	VPGC	VPGC	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
TOA and Surface Radiation Flux Products																	
CER06 [SRBAVG]	T&A1	3				VPGI	VPGC	VPGC	VFGI	VFGI	VFGI	VFGC	VFGC	VFGC	VFGC	AFGC	AFGC
LIS																	
LIS01	TRMM	1A	VFII	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
LIS02	TRMM	1B	VFII	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
LIS03	TRMM	1B	VFII	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
LIS04	TRMM	1B	VFII	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
LIS05	TRMM	2	VFII	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
LIS06	TRMM	2	VFII	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
LIS07	TRMM	2	VFII	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
LIS08	TRMM	3	VFII	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
LIS09	TRMM	3	VFII	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
LIS10	TRMM	4	VFII	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MISR																	
MIS01	AM1	1A					VFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MIS02	AM1	1B1					VFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MIS03	AM1	1B2					VFRC	AFRC	AFIC	AFIC	AFIC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MIS04	AM1	2					VPRC	VFRC	SFIC	AFIC	AFIC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MIS05	AM1	2					VFRC	VFRC	SFIC	SFIC	AFIC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MIS06 (Combined with MIS05, i.e. combined Aerosol & Surface product)																	
MIS07	AM1	3									AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MIS08	AM1	3									AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MIS09	AM1	3									AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MIS10	AM1	1B2					VFGX	AFGX	AFGX	AFGX	AFGX	AFGX	AFGX	AFGX	AFGX	AFGX	AFGX
MIS11	AM1	1B1					VFXX	AFXX	AFXX	AFXX	AFXX	AFXX	AFXX	AFXX	AFXX	AFXX	AFXX
MIS12	AM1	2					VFGX	VFGX	AFGX	AFGX	AFGX	AFGX	AFGX	AFGX	AFGX	AFGX	AFGX
MODIS																	
Product ID	Platform	Prod. Level	3rd-97	4th-97	1st-98	2nd-98	3rd-98	4th-98	1st-99	2nd-99	3rd-99	4th-99	1st-00	2nd-00	3rd-00	4th-00	1st-01

MOD01	AM1	1A	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD02	AM1	1B	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD03	AM1	1A	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD04	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD05	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD06	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD07	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD08	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD09	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD10	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD11	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD12	AM1	3	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD13	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD14	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD15	AM1	4	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD16	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD17	AM1	4	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD18	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD19	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD20	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD21	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD22	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD23	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD24	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD25	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD26	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD27	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD28	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD29	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD30	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD31	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD32	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD33	AM1	3	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD34	AM1	3	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD35	AM1	1B	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD36	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD37	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD38	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD39	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD40	AM1	3	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD41	AM1	3	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD42	AM1	3	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOD43	AM1	2	VFGG	VFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG	AFGG
MOPITT													
MOP01	AM1	1B	VFGC	VFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MOP02	AM1	2	VPGI	VFGI	VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MOP03	AM1	2	VPGI	VFGI	VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MOP04	AM1	2	VPGI	VFGI	VFGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MOP05	AM1	3			VPGI	VPGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MOP06	AM1	3			VPGI	VPGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC
MOP07	AM1	3			VPGI	VPGI	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC	AFGC

EXHIBIT B-1: AHWGP Product Availability Scenarios

B.2 MOPITT PRODUCTION PROCESSING

We have compared various modeling parameters including processing load, input/output files sizes, external data dependence, AHWGP product availability scenarios for the MOPITT standard products from different sources including the ATBD, Excel spreadsheet data in the PDR Technical Baseline, and the information in the AHWGP Gopher server. We chose MOPITT for this comparison because status of modeling efforts for this instrument was more mature than the other instruments The following table contains the relevant MOPITT product processing information available from different sources. The comparison points to some discrepancies that need to be resolved before they are input into the Production Model.

Product Name (ID)	ATBD #	AHWGP Product Availability Time frame (Quarter/Year)	Process (ID)	Total No of MFPO per Execution	Processing Power (MFLOPS)	No. of Exec. per Day	Input Files ID (Size in MB)	Time Cover (min)	No. of Read per Exec.	Output File ID (Size in MB)	Time Cover (min)	No. of Write per Exec	Comment
Level 1B Radiance (MOP01)	ATBD-MOP01 (Retrieval of Carbon)	VFGC 3/98 VFGC 4/98 AFGC (3/98-4/02)			*0.160 (ATBD)*								Epochs PROCTIM XLS fghijk...., ghijk...in PROCESS LS (f+2/98; g=3/98)
			Level 1 Processing (MOPL1)	16,800	0.19	1	MOP-00 (255.24) MOP-CH (1) MOP-1P1 (1)	1440 43,200 0	1 1 1	MOP 01 (101) MOP-01D (255.24) MOP-CH	1440 0	1 1 1	
		3/98	Level 1 QA and Error Analysis (MOPL1Q1-C)	1350	0.02	1	MOP-01 MOP-01D		1 1	MOP01Q-C (356.24)		1	
		4/98-4/02	Level 1 QA and Error Analysis (MOPL1Q1-D)	900	0.01	1	MOP-01 MOP01-D		1 1	MOP01Q-D (10)	0	1	
Level 2 Processing (CH4 Burden-Total Column) (MOP02)	ATBD-MOP-02 (Retrieval of Carbon)	VPGI 2/98 VFGI (3/98-1/99) AFGC (2/99-4/02)			*2.780 (ATBD)*								Time li do not ma with th from AHWGP
		3/98-1/99	Level 2 Processing (MOPL2-C)	1,112,775	3.22	0.25	MOP-01 MOP-1P2 (1) MOP-AX (50) **ANC_EDC_DEM (200)** **ANC_NMC_PROF (3)** **ANC_NMC_SURF (3)** MOP-AC MOP-SurfP	0 0 0 720 720	1 1 1 1 4 4 1 1 1 1	MOP-02 (0.02) MOP-02D MOP-02B (0.00)		1 1 1	
		3/98-1/99	Level 2 QA and Error Analysis (MOPL2-Qi-C)	1,350	0.02	1	MOP-OC MOP-NC		1 1	MOP02Q-C		1	
		2/99-3/00	Level 2 Processing	1,502,250	17.39	1	MOP-02 MOP-02D		1 1 1	MOP-02 (0.07) MOP-02D MOP-02B (0.01)		1 1 1	

Product Name (ID)	ATBD #	AHWGP Product Availability Time frame (Quarter/Year)	Process (ID)	Total No of MFPO per Execution	Processing Power (MFLOPS)	No. of Exec. per Day	Input Files ID (Size in MB)	Time Cover (min)	No. of Read per Exec.	Output File ID (Size in MB)	Time Cover (min)	No. of Write per Exec	Commen
			(MOPL2-E)				MOP-01 (0.01)		1				
							MOP-1P2		4				
							MOP-AX		585				
							ANC_EDC_DEM		4				
							ANC_NMC_PROF		585				
							MOD30-L2-G		1				
							ANC-NMC-SURF		1				
							MOD06-L2-G		1				
							MOP-AC		1				
							MOP-SurfP						
		2/99-4/02		1350		1	MOP-OC		1	MOP02Q-D		1	
			Level 2 QA (MOPL2Qi-D)		0.02		MOP-NC		1				
							MOP-02						
		4/00-4/02		1,502,250		1	MOP-02D		1	MOP-02 (0.07)		1	
			Level 2 Processing (MOPL2-H)		17.39				1	MOP-02D		1	
									1	MOP-02B (0.01)		1	
							MOP-01		1				
							MOP-1P2		4				
							MOP-AX		585				
							ANC_EDC_DEM		4				
							ANC_NMC_PROF		585				
							MOD30_L2_G		1				
							ANC_NMC_SURF		1				
							MOD06_L2_G		1				
							AIRS-05		1				
							AIRS-07		1				
							MOP-AC		1				
							MOP-SurfP						
							MOP-OC						
							MOP-NC						
Level 3 Processing (CO Profiles) (MOP03)	ATBD-MOP-02 (Retrieval of Carbon)	VPGI 2/98 VFGI (3/98-1/99) AFGC (2/99-4/02)			*7.300 (ATBD)*								Time li do not ma with th from AHWGP
		1/99-4/02	Level 3 Processing (MOPl3)	28,290	0.058	0.142	MOP-02		7	MOP-03 (0.00)		1	
										MOP-03D		1	
										MOP-03B (0.00)		1	
		1/99-2/99	Level 3 QA and	900		0.142	MOP-03		1	MOPL3Qi-E		1	

Product Name (ID)	ATBD #	AHWGP Product Availability Time frame (Quarter/Year)	Process (ID)	Total No of MFPO per Execution	Processing Power (MFLOPS)	No. of Exec. per Day	Input Files ID (Size in MB)	Time Cover (min)	No. of Read per Exec.	Output File ID (Size in MB)	Time Cover (min)	No. of Write per Exec	Comment
			Error Analysis (MOPL3Qi-E)		00		MOP-03D		1				
		3/99-4/02	Level 3 QA and Error Analysis (MOPL3Qi-F)	900	00	0.142	MOP-03 MOP-03D		1 1	MOPL3Qi-F		1	

* N Standard product not available, V Undergoing validation, users beware, A Available for general use, P Partial parameter generation,
F Full parameter generation
R Regional coverage, such as for initial algorithm proving, I Intermediate, moderately regular coverage, G Global coverage
S Sporadic: only a few, irregular times in a month, I Intermittent: regular, Moderately frequent sampling, C Continuous: Large fraction of possible samples taken
X Designation not applicable for that field

EXHIBIT B-1: MOPITT Product Processing

B.3 BONeS Input File (F_Desc.txt) Content for ASTER

We obtained a printout of the input file F_Desc.txt from the version of BONeS model the IV&V team received on January 26, 1995. The file contains some of the data that HAIS has received from the AHWGP for input into BONeS to simulate ASTER production processing. We have used this data for comparison with those in the Excel spreadsheet in the PDR Technical Baseline and the ASTER Product Processing Diagram (Exhibit 4-5) supplied by the Instrument Team. We have chosen ASTER to do this comparison because only ASTER data was available in this version of the model

Numerical codes used:

File ID:
Unique number designating a particular file is shown below.

AST_ANC_01	0	AST_10	1	AST_ANC_02	2	ANC_DEM	3
ANC_LAND_COVER	4	AST_L1B	5	ANC_LAND_SEA	6	AST_06B	7
AST_06C	8	AST_DS_TMP1	9	AST_ANC_06A	10	AST_DS_TMP2	11
AST_ANC-03	12	AST_04	13	AST_BT_TMP1	14	AST_ANC_MIS05	15
AST_09B	16	AST_ANC_MIS12	17	AST_09A	18	AST_07B	19
AST_ANC_04	20	AST_07A	21	AST_NMC	22	AST_ANC_05	23
AST_NMC_TOMS	24	AST_ANC_MOD30	25	AST_DEM_GRD_TEMP1	26	AST_DEM_PIX_TMP1	27
AST_MODTN1	28	AST_MODTN2	29	AST_09C	30	AST_ANC_MOD30B	31
AST_ANC_06	32	AST_05_08	33	AST_T/E_TMP	34	AST_11	35
AST_12	36	AST_PVI/SBI_TMP	37	AST_13	38	ANC_ECOSYS_DB	39
AST_ANC_07	40	AST_ANC_MOD10	41	AST_14	42	AST_DEM_GRD_TMP2	43
AST_DEM_PIX_TMP2	44	AST_MODTN4	45	AST_MODTN4	46		

Instrument ID:

0 = Other	1 = CERES (TRMM)	2 = VIRS (TRMM)	3 = PR (TRMM),
4 = TMI (TRMM)	5 = GV (TRMM)	6 = LIS (TRMM)	7 = ASTER (AM-1),
8 = CERES (AM-1)	9 = MISR (AM-1)	10 = MODIS (AM-1)	11 = MOPITT (AM-1)

Archive Site:

0 = Other	1 = ASF	1 = EDC	3 = GSFC
4 = JPL	5 = LaRC	6 = MSFC	7 = NSIDC

File Disposition:

0 = Other	1 = Archive	2 = Interim	3 = Permanent
4 = Temporary	5 = Transfer to SCF		

External/Root Flag:
0 = Not Applicable 1 = Level 0 data 2 = External data

Ingest Media Flag:
0 = Electronic 1 = Physical media

# of Files on Media:
0 = Default n = # of files to read from media

Source:

0 = Unknown	1 = ASF	2 = EDC	3 = GSFC
4 = JPL	5 = LaRC	6 = MSFC	7 = NSIDC
8 = TSDIS	9 = PACOR	10 = EDOS	11 = Japan GS (ASTER)
12 = Landsat-7 GS	13 = NMC		

Contents of the File

File ID	Inst. ID	Arch. Site	File Disp.	File Size (MB)	Temporal Coverage (min.)	External/ Root Flag	Ingest Media Flag	# of Files	Source
0	7	2	3	0.5	500000	0	0	0	2
1	7	2	1	18	7.91208791	0	0	0	2
2	7	2	3	1	500000	0	0	0	2
3	0	2	3	200	500000	0	0	0	2
4	0	2	3	250	500000	0	0	0	2
5	7	2	1	44226	1440	1	1	182	11
6	0	2	3	0.15	500000	0	0	0	2
7	7	2	1	13.2	7.91208791	0	0	0	2
8	7	2	1	1.1	7.91208791	0	0	0	2
9	7	2	4	16	7.91208791	0	0	0	2
10	7	2	1	52.8	7.91208791	0	0	0	2
11	7	2	4	16	7.91208791	0	0	0	2
12	0	2	3	0.3	500000	0	0	0	2
13	7	2	1	6	20.5714286	0	0	0	2
14	7	2	4	0.64	7.91208791	0	0	0	2
15	9	5	2	500	98	0	0	0	5
16	7	2	1	61.6	20.5714286	0	0	0	2
17	9	5	2	1.6	500000	0	0	0	5
18	7	2	1	176	20.5714286	0	0	0	2
19	7	2	1	62	20.5714286	0	0	0	2
20	0	2	3	10000	500000	0	0	0	2
21	7	2	1	176	20.5714286	0	0	0	2
22	0	3	2	2	180	0	0	0	3
23	0	2	3	1	500000	0	0	0	2
24	0	3	2	0.5	1440	0	0	0	3
25	0	3	2	35.8	2.46153846	0	0	0	3
26	7	2	4	0.03	7.91208791	0	0	0	2
27	7	2	4	128	7.91208791	0	0	0	2
28	7	2	4	2	7.91208791	0	0	0	2
29	7	2	4	0.226	7.91208791	0	0	0	2
30	7	2	1	6	20.5714286	0	0	0	2
31	0	3	2	35.8	2.46153846	0	0	0	3
32	0	2	3	0.001	500000	0	0	0	2

File ID	Inst. ID	Arch. Site	File Disp.	File Size (MB)	Temporal Coverage (min.)	External/ Root Flag	Ingest Media Flag	# of Files	Source
33	7	2	1	8	20.5714286	0	0	0	2
34	7	2	4	3.6	7.91208791	0	0	0	2
35	7	2	1	18	51.4285714	0	0	0	2
36	7	2	1	36	51.4285714	0	0	0	2
37	7	2	4	18	7.91208791	0	0	0	2
38	7	2	1	18	240	0	0	0	2
39	0	2	3	15	500000	0	0	0	2
40	0	2	2	2	10080	0	0	0	0
41	0	7	2	7.2	10080	0	0	0	7
42	7	2	1	35	1440	0	0	0	2
43	7	2	4	0.03	7.91208791	0	0	0	2
44	7	2	4	128	7.91208791	0	0	0	2
45	7	2	4	2	7.91208791	0	0	0	2
46	7	2	4	0.226	7.91208791	0	0	0	2

EXHIBIT B-3: BONeS Input File (L_Desc.txt) Content For ASTER

APPENDIX C: PERFORMANCE MODEL ANALYSIS DETAIL

The detailed set of Exhibits shown in Appendix C will provide the basis for the evaluation of the model's completeness and correctness, with respect to the ECS design representation, when the final version of the model and associated documentation become available. The Exhibits are TBD for this release of the TAR

This Exhibit accounts for subsystem distribution by DAAC site.

Site	Client	Inter-operability	Data Mgt	Data Server	Ingest	Data Processing	Planning
ASF							
CIESIN							
EDC							
GSFC							
JPL							
LaRC							
MSFC							
NSIDC							

EXHIBIT C-1: Subsystem Sites

This Exhibit will account for subsystem services as defined in the SDS.

Subsystem	Service Class	Model Coverage
Client	Desktop	
	Scientist Workbench	
	UI Mgt & Presentation	
	Operating Support	
Interoperability	CSMS	
	Advertisement	
	Subscription	
Data Management	DIM	
	LIM	
	Data Dictionary	
Data Server	Data Server	
	Data Type	
	Data Storage & Management	
	Data Distribution	
	Administration	
	Schema Generation	
Ingest	Ingest Client	
Planning	Production Management	
	Production Planning	
Data Processing	Process Management	
	Process Queue	
	Process Execution	
	Executable Process	
	Resource Management	
	Process Integration & Test	

EXHIBIT C-2: Subsystem Services

This Exhibit will account for Data Management Subsystem activities and interfaces as defined in the SDS.

Interface	Activity/Process	Model Coverage
Interoperability	Notify-input	
	Advertise-output	
	Subscription-output	
Data Server	Results Set-input	
	SessionMgtResp-input	
	Schema-input	
	Data Dictionary-input	
	SessionMgtReq-output	
	Search Req-output	
	Access Req-output	
V0	Product Req-input	
	Inventory-input	
	Browse-input	
	Guide-input	
	Status-input	
Client	Search Req-input	
	Access Req-input	
	SessionMgtReq-input	
	Subscriptions-input	
	Results Set-output	
	SessionMgtResp-output	
	Notification-output	

EXHIBIT C-3: Data Management Subsystem Activities/Processes

This Exhibit will account for Data Processing Subsystem activities and interfaces as defined in the SDS.

Interface	Activity/Process	Model Coverage
Data Server	Processing Info-output	
	Processing Status-output	
	Processing Results-output	
	Access Req-output	
	Results Ref-output	
	Completion Notify-output	
	Processing Info Req-input	
	Processing Control Req-input	
	Processing DatC-input	
Planning	Processing Req-input	
	Processing Info Req-input	
	Processing Control Req-input	
	Processing Info-output	
	Completion Notify-output	
	Results Ref-output	
MSS/LSM	Resource Info-input	
	Maintenance Schedule-input	
MSS/SMC	Processing Info Req-input	
	Processing Status Req-input	
	Processing Status-output	
	Processing Info-output	
Interoperability	Service Advertise-output	
	Subscriptions-output	
	Notification-input	
Client	Processing Info Req-input	
	Processing Control Req-input	
	Processing Info-output	
	Processing Status-output	

EXHIBIT C-4: Data Processing Subsystem Activities/Processes

This Exhibit will account for Planning Subsystem activities and interfaces as defined in the SDS.

Interface	Activity/Process	Model Coverage
Data Server	Data Available Notify-input	
	Data Available Schedule-input	
	Processing Req-input	
	Subscription-input	
	Search Req-output	
	Subscription-output	
	Completion Notify-output	
	Results Reference-output	
	Production Schedule-output	
Data Processing	Processing Req-output	
	Processing Info Req-output	
	Processing Control Req-output	
	Processing Status-input	
	Results Reference-input	
	Completion Notify-input	
MSS/LSM	Resource Info-input	
	Maintenance Schedule-input	
	Plans-output	
MSS/SMC	Planning Info-output	
	Plans-output	
	Plans-input	
Interoperability	Service Advertise-output	
	Subscriptions-output	
	Notification-input	
Client	Plan Info-output	
	Plan Status-output	
	Change Notify-output	
	Plan Control Req-input	
	Plan Info Req-input	
	Subscription-input	

EXHIBIT C-5: Planning Subsystem Activities/Processes

This Exhibit will account for Data Server Subsystem activities and interfaces as defined in the SDS.

Interface	Activity/Process	Model Coverage
MSS/SMC	Status Req-input	
	Log Req-input	
	Status Log-output	
Other DAACs	Data Group 2-output	
	Algorithms-output	
ADCs/ODCs	Data Group 1-output	
	Algorithms-output	
	Product Req-input	
Ingest	Ingested DatC-input	
V0	Inventory-output	
	Browse DatC-output	
	Guide-output	
	Product Req-output	
	Dependent Valids-output	
Planning	Processing Req-output	
	Subscriptions-output	
	Data Available Schedules-output	
	Subscriptions-output	
	Search Results-output	
	Results Reference-output	
	Completion Notify-output	
	Production Schedule-input	
	Notifications-input	
	Search Req-input	
IPs	Data Group 2-output	
	DARs-output	
	DAR Status-input	
Data Processing	Processing DatC-output	
	Processing Status-input	
	Processing Results-input	
	Access Req-input	
	Results Reference-input	
	Completion Notify-input	
Interoperability	Service Advertise-output	
	Subscriptions-output	
	Notification-input	
FOS	Historic DatC-input	

Interface	Activity/Process	Model Coverage
	DAR Status Updates-input	
	QL Images-output	
	Historic DatC-output	
	DAR-output	
SCF	QA Req-input	
	Data Group 4-output	
Data Management	Search Results-input	
	Access Req-input	
	Session Mgt Req-input	
	Search Results-output	
	SchemC-output	
	Data Dictionary-output	
	Session Mgt Resp-output	
Client	Search Results-input	
	Access Req-input	
	Admin Req-input	
	Search Results-output	
	Data Types-output	
TRMM (TSDIS)	Data Group 1-output	
	Algorithms-output	
	Product Req-input	
Users	Data Group 1-output	
	DAR Status-output	
	Status-output	

EXHIBIT C-6: Data Server Subsystem Activities/Processes

This Exhibit will account for Interoperability Subsystem activities and interfaces as defined in the SDS.

Interface	Activity/Process	Model Coverage
Ingest	Notification-output	
	Advertisement-input	
	Subscription-input	
Planning	Notification-output	
	Advertisement-input	
	Subscription-input	
Data Processing	Notification-output	
	Advertisement-input	
	Subscription-input	
Data Server	Notification-output	
	Advertisement-input	
	Subscription-input	
Data Management	Notification-output	
	Advertisement-input	
	Subscription-input	
Client	Notification-output	
	Search Result-output	
	Access Result-output	
	Advertisement Info-output	
	Subscription-input	
	Access Request-input	
	Search Request-input	

EXHIBIT C-7: Inoperability Subsystem Activities/Processes

This Exhibit will account for Ingest Subsystem activities and interfaces as defined in the SDS.

Interface	Activity/Process	Model Coverage
Land Sat-7	Data Group 1-input	
ADCs/ODCs	Data Group 1-input	
	Algorithms-input	
	Product Requests-input	
SDPF	L0 DatC-input	
	Orbit/Attitude DatC-input	
EDOS	L0 DatC-input	
	Orbit/Attitude DatC-input	
	Quick Look DatC-input	
	Data Avaibility Schedule-input	
	Backup Data Request-input	
Users	User Methods-input	
V0	Migration DatC-input	
MSS/SMC	Ingest Status Requests-input	
	Ingest Log Requests-input	
	Ingest Status-output	
	Ingest Log-output	
Client	Ingest Status Requests-input	
	Ingest Log Requests-input	
	Ingest Control Requests-input	
	Ingest Status-output	
	Ingest Log-output	
Data Server	Ingested DatC-output	
Interoperability	Notification- input	
	Advertisement- output	
	Subscription- output	
Other DAACs	Data Group 1-input	
	Algorithms-input	
FDF	Refined/Repeated O/A Req-output	
	Orbit/Attitude DatC-input	
	Predicted Orbit DatC-input	
IPs	Data Group 2-input	
	Algorithms-input	
SCFs	Data Group 1-input	
	Algorithms-input	
TRMM (TSDIS)	Data Group 1-input	
	Algorithms-input	

EXHIBIT C-8: Ingest Subsystem Activities/Processes

This Exhibit will account for Client Subsystem activities and interfaces as defined in the SDS.

Interface	Activity/Process	Model Coverage
SMC	User Registration Status-output	
	Status Request-input	
Data Servers	DAR-output	
	Search Requests-output	
	Access Requests-output	
	Session Mgt Requests-output	
	Subscriptions-output	
	Results Set-input	
	Session Mgt Reponses-output	
	Notification-output	
Data Management	Search Requests-output	
	Access Requests-output	
	Session Mgt Requests-output	
	Subscriptions-output	
	Results Set-input	
	Session Mgt Reponses-output	
	Notification-output	
Ingest	Ingest Status Request-output	
	Ingest Log Request-output	
	Ingest Control Request-output	
	Ingest Status-input	
	Ingest Log-input	
Processing	Processing Info Request-output	
	Processing Control Request-output	
	Processing Info-input	
	Processing Status-input	
Users	User Registration-input	
	Search Requests-input	
	Access Requests-input	
	DARS-input	
	Results Set-output	
	DAR Status-output	
Planning	Plan Info-input	
	Plan Status-input	
	Change Notification-input	
	Plan Control Requests-output	
	Plan Info Requests-output	

Interface	Activity/Process	Model Coverage
	Subscriptions- input	
Interoperability	Notification- input	
	Search Result- input	
	Access Result- input	
	Advertisement Info- input	
	Subscription- output	
	Access Request- output	
	Search Request- output	
Local DAAC	Mgt & Ops Status-output	
	Mgt & Ops Cmd-input	

EXHIBIT C-9: Client Subsystem Activities/Processes

Files are TBD.

Files	Archive Site	Epochs		Model Coverage
		Start	End	
ACR01				

EXHIBIT C-10: ACRIM Files

Files	Archive Site	Epochs		Model Coverage
		Start	End	
AIR01				
AIR02				
AIR03				
AIR04				
AIR05				
AIR06				
AIR07				
AIR08				
AIR09				
AIR10				

EXHIBIT C-11: AIRS Files

Files are TBD.

Files	Archive Site	Epochs Start	Epochs End	Model Coverage

EXHIBIT C-12: AMSU Files

Files	Archive Site	Epochs Start	Epochs End	Model Coverage
AST_L1A	EDC	3Q 98	4Q 02	
AST_L1B	EDC	3Q 98	4Q 02	
AST_L10	EDC	3Q 98	4Q 02	
AST_06B	EDC	3Q 98	4Q 02	
AST_06C	EDC	3Q 98	4Q 02	
AST_06A	EDC	3Q 98	4Q 02	
AST_04	EDC	3Q 98	4Q 02	
AST_09B	EDC	3Q 98	4Q 02	
AST_09A	EDC	3Q 98	4Q 02	
AST_07B	EDC	3Q 98	4Q 02	
AST_07A	EDC	3Q 98	4Q 02	
AST_09C	EDC	3Q 98	4Q 02	
AST_05_08	EDC	3Q 98	4Q 02	
AST_11	EDC	3Q 98	4Q 02	
AST_12	EDC	3Q 98	4Q 02	
AST_13	EDC	3Q 98	4Q 02	
AST_14	EDC	3Q 98	4Q 02	

EXHIBIT C-13: ASTER Files

Files are TBD.

Files	Archive Site	Epochs Start	Epochs End	Model Coverage

EXHIBIT C-14: AVHRR Files

Files	Archive Site	Epochs		Model Coverage
		Start	End	
CER01T	LaRC	3Q 97	4Q 00	
CER01A1	LaRC	3Q 98	2Q 03	
CER01P1	LaRC	1Q 01	2Q 03	
CER02T	LaRC	3Q 97	4Q 00	
CER02A	LaRC	3Q 98	2Q 03	
CER02P	LaRC	1Q 01	2Q 03	
CER03aT	LaRC	3Q 97	4Q 00	
CER13aT	LaRC	3Q 97	4Q 00	
CER14aT	LaRC	3Q 97	4Q 00	
CER03aA	LaRC	3Q 98	3Q 01	
CER13aA	LaRC	3Q 98	3Q 01	
CER14aA	LaRC	3Q 98	3Q 01	
CER03aP	LaRC	1Q 01	3Q 01	
CER13aP	LaRC	1Q 01	3Q 01	
CER14aP	LaRC	1Q 01	3Q 01	
CER03bTA	LaRC	3Q 98	3Q 00	
CER13bTA	LaRC	3Q 98	3Q 00	
CER14bTA	LaRC	3Q 98	3Q 00	
CER03bAP	LaRC	1Q 01	3Q 01	
CER13bAP	LaRC	1Q 01	3Q 01	
CER14bAP	LaRC	1Q 01	3Q 01	
CER11T	LaRC	3Q 97	4Q 97	
CER11T	LaRC	3Q 97	3Q 00	
CER11A1	LaRC	3Q 98	4Q 98	
CER11A1	LaRC	3Q 98	3Q 01	
CER11P1	LaRC	1Q 01	2Q 01	
CER11P1	LaRC	1Q 01	3Q 01	
CER04aT	LaRC	4Q 97	4Q 00	
CER04aA	LaRC	3Q 98	4Q 00	
CER04bP	LaRC	1Q 01	2Q 01	
CER04aT	LaRC	1Q 98	4Q 00	
CER04aA	LaRC	3Q 98	4Q 00	
CER04bP	LaRC	3Q 01	2Q 03	
CER04bA	LaRC	1Q 01	2Q 03	
CER05aT	LaRC	1Q 98	4Q 00	
CER05aA	LaRC	3Q 98	2Q 03	
CER05bA	LaRC	1Q 01	2Q 03	
CER05bP	LaRC	1Q 01	2Q 03	
CER07aT	LaRC	1Q 98	4Q 00	
CER07aA	LaRC	3Q 98	4Q 00	
CER07bA	LaRC	1Q 01	2Q 03	

Files	Archive Site	Epochs		Model Coverage
		Start	End	
CER07bP	LaRC	1Q 01	2Q 03	
CER07CTA	LaRC	3Q 98	4Q 00	
CER07dAP	LaRC	1Q 01	2Q 03	
CER08aT	LaRC	1Q 98	3Q 00	
CER15aT	LaRC	1Q 98	4Q 00	
CER08aA	LaRC	3Q 98	4Q 00	
CER15aA	LaRC	3Q 98	4Q 00	
CER08bA	LaRC	1Q 01	2Q 03	
CER15bA	LaRC	1Q 01	2Q 03	
CER08bP	LaRC	1Q 01	2Q 03	
CER15bP	LaRC	1Q 01	2Q 03	
CER08CTA	LaRC	3Q 98	4Q 00	
CER15CTA	LaRC	3Q 98	4Q 00	
CER08dAP	LaRC	1Q 01	2Q 03	
CER15dAP	LaRC	1Q 01	2Q 03	
CER12T	LaRC	3Q 97	4Q 97	
CER12T	LaRC	3Q 97	4Q 00	
CER12A	LaRC	3Q 98	4Q 98	
CER12A	LaRC	3Q 98	2Q 03	
CER12P	LaRC	1Q 01	2Q 01	
CER12P	LaRC	1Q 01	2Q 03	
CER06AT	LaRC	3Q 97	4Q 00	
CER06aA	LaRC	3Q 98	2Q 03	
CER06aP	LaRC	1Q 01	2Q 03	
CER06bTA	LaRC	3Q 98	4Q 00	
CER06bAP	LaRC	1Q 01	2Q 03	
CERX01T	LaRC	3Q 97	4Q 00	
CERX01A	LaRC	3Q 98	2Q 03	
CERX01P	LaRC	3Q 98	2Q 03	
CERX02T	LaRC	3Q 97	4Q 00	
CERX02A	LaRC	3Q 98	2Q 03	
CERX02P	LaRC	2Q 01	2Q 03	
CERX06	LaRC	3Q 97	4Q 97	
CERX06	LaRC	3Q 97	2Q 03	

EXHIBIT C-15: CERES Files

Files are TBD.

Files	Archive Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-16: DORIS Files

Files	Archive Site	Epochs		Model Coverage
		Start	End	
EOSP01				
EOSP02				
EOSP03				
EOSP04				
EOSP05				
EOSP06				
EOSP07				
EOSP08				

EXHIBIT C-17: ESOP Files

Files are TBD.

Files	Archive Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-18: ETM Files

Files	Archive Site	Epochs Start End	Model Coverage
GLA01			
GLA02			
GLA03			
GLA04			
GLA05			
GLA06			
GLA07			
GLA08			
GLA09			
GLA10			
GLA11			

EXHIBIT C-19: GLAS Files

Files	Archive Site	Epochs Start End	Model Coverage
HIR01			
HIR02			
HIR03			
HIR04			
HIR05			
HIR06			
HIR07			
HIR08			
HIR09			
HIR10			
HIR11			
HIR12			
HIR13			
HIR14			
HIR15			
HIR16			
HIR17			
HIR18			
HIR19			
HIR20			
HIR21			
HIR22			
HIR23			
HIR24			

Files	Archive Site	Epochs		Model Coverage
		Start	End	
HIR25				
HIR26				
HIR27				
HIR28				
HIR29				

EXHIBIT C-20: HIRDLS Files

Files	Archive Site	Epochs		Model Coverage
		Start	End	
LIS00	MSFC	3Q 97	2Q 03	
LISCAL	MSFC	3Q 97	2Q 03	
LIS02	MSFC	3Q 97	2Q 03	
LIS03	MSFC	3Q 97	2Q 03	
LIS04	MSFC	3Q 97	2Q 03	
LIS05	MSFC	3Q 97	2Q 03	
LIS06	MSFC	3Q 97	2Q 03	
LIS07	MSFC	3Q 97	2Q 03	
LIS08	MSFC	3Q 97	2Q 03	
LIS07B	MSFC	3Q 97	2Q 03	
LIS08B	MSFC	3Q 97	2Q 03	
LIS09	MSFC	3Q 97	2Q 03	
LIS10	MSFC	3Q 97	2Q 03	

EXHIBIT C-21: LIS Files

Files are TBD.

Files	Archive Site	Epochs		Model Coverage
		Start	End	
MHS03				
MHS02				

EXHIBIT C-22: MHS Files

Files	Archive Site	Epochs Start End	Model Coverage
MIM01			
MIM02			
MIM03			
MIM04			
MIM05			
MIM06			
MIM07			
MIM08			
MIM09			
MIM10			
MIM11			
MIM12			
MIM13			
MIM14			
MIM15			
MIM16			
MIM17			
MIM18			
MIM19			
MIM20			
MIM21			
MIM22			
MIM23			
MIM24			
MIM25			
MIM26			
MIM27			

EXHIBIT C-23: MIMR Files

Files	Archive Site	Epochs Start	End	Model Coverage
MIS-1ASx	LaRC	3Q 98	2Q 03	
MIS-1AE	LaRC	3Q 98	2Q 03	
MIS-1ACx	LaRC	3Q 98	2Q 03	
MIS-1ACE	LaRC	3Q 98	2Q 03	
MIS-1-BW	LaRC	3Q 98	2Q 03	
MIS-1AN	LaRC	3Q 98	2Q 03	
MIS-1B2Sx	LaRC	3Q 98	3Q 98	
MIS-1B2Hx	LaRC	3Q 98	3Q 98	
MIS-1B2SS	LaRC	3Q 98	3Q 98	
MIS-1B2SH	LaRC	3Q 98	3Q 98	
MIS1B2Sx	LaRC	4Q 98	4Q 98	
MIS-1B2Hx	LaRC	4Q 98	4Q 98	
MIS-1B2SS	LaRC	4Q 98	4Q 98	
MIS-1B2SH	LaRC	4Q 98	4Q 98	
MIS-1B2Sx	LaRC	1Q 99	1Q 99	
MIS-1B2Hx	LaRC	1Q 99	1Q 99	
MIS-1B2SS	LaRC	1Q 99	1Q 99	
MIS-1B2SH	LaRC	1Q 99	1Q 99	
MIS-1B2Sx	LaRC	2Q 99	2Q 99	
MIS-1B2Hx	LaRC	2Q 99	2Q 99	
MIS-1B2SS	LaRC	2Q 99	2Q 99	
MIS-1B2SH	LaRC	2Q 99	2Q 99	
MIS-1B2Sx	LaRC	3Q 99	3Q 99	
MIS-1B2Hx	LaRC	3Q 99	3Q 99	
MIS-1B2SS	LaRC	3Q 99	3Q 99	
MIS-1B2SH	LaRC	3Q 99	3Q 99	
MIS-1B2Sx	LaRC	1Q 00	2Q 03	
MIS-1B2Hx	LaRC	1Q 00	2Q 03	
MIS-1B2SS	LaRC	1Q 00	2Q 03	
MIS-1B2SH	LaRC	1Q 00	2Q 03	
MIS-2TC	LaRC	3Q 98	3Q 98	
MIS-2TCS	LaRC	3Q 98	3Q 98	
MIS-2TC	LaRC	4Q 98	4Q 98	
MIS-2TCS	LaRC	4Q 98	4Q 98	
MIS-2TC	LaRC	1Q 99	1Q 99	
MIS-2TCS	LaRC	1Q 99	1Q 99	
MIS-2TC	LaRC	2Q 99	2Q 99	
MIS-2TCS	LaRC	2Q 99	2Q 99	
MIS-2TC	LaRC	3Q 99	3Q 99	
MIS-2TCS	LaRC	3Q 99	3Q 99	

Files	Archive Site	Epochs		Model Coverage
		Start	End	
MIS-2TC	LaRC	4Q 99	2Q 03	
MIS-2TCS	LaRC	4Q 99	2Q 03	
MIS-2AS	LaRC	3Q 98	3Q 98	
MIS-2ASS	LaRC	3Q 98	3Q 98	
MIS-2-BW1	LaRC	3Q 98	3Q 98	
MIS-2-BW2	LaRC	3Q 98	3Q 98	
MIS-2AS	LaRC	4Q 98	4Q 98	
MIS-2ASS	LaRC	4Q 98	4Q 98	
MIS-2-BW1	LaRC	4Q 98	4Q 98	
MIS-2-BW2	LaRC	4Q 98	4Q 98	
MIS-2AS	LaRC	1Q 99	1Q 99	
MIS-2ASS	LaRC	1Q 99	1Q 99	
MIS-2-BW1	LaRC	1Q 99	1Q 99	
MIS-2-BW2	LaRC	1Q 99	1Q 99	
MIS-2AS	LaRC	2Q 99	2Q 99	
MIS-2ASS	LaRC	2Q 99	2Q 99	
MIS-2-BW1	LaRC	2Q 99	2Q 99	
MIS-2-BW2	LaRC	2Q 99	2Q 99	
MIS-2AS	LaRC	3Q 99	3Q 99	
MIS-2ASS	LaRC	3Q 99	3Q 99	
MIS-2-BW1	LaRC	3Q 99	3Q 99	
MIS-2-BW2	LaRC	3Q 99	3Q 99	
MIS-2AS	LaRC	4Q 99	2Q 03	
MIS-2ASS	LaRC	4Q 99	2Q 03	
MIS-2-BW1	LaRC	4Q 99	2Q 03	
MIS-2-BW2	LaRC	4Q 99	2Q 03	

EXHIBIT C-24: MISR Files

Files	Archive Site	Epochs Start End	Model Coverage
MLS01			
MLS02			
MLS03			
MLS04			
MLS05			
MLS06			
MLS07			
MLS08			
MLS09			
MLS10			
MLS11			
MLS12			
MLS13			
MLS14			
MLS15			
MLS16			
MLS17			
MLS18			

EXHIBIT C-25: MLS Files

Files	Archive Site	Epochs Start End	Model Coverage
MOD01_L1A_G	GSFC	3Q 98 3Q 02	
MOD02_L1B_G	GSFC	3Q 98 3Q 02	
MOD03_L1A_G	GSFC	3Q 98 3Q 02	
MOD03_L1A_QADATA	GSFC	3Q 98 3Q 02	
MOD4_L3_DY	GSFC	3Q 98 3Q 02	
MOD4_L3_WK	GSFC	3Q 98 4Q 02	
MOD4_L3_MN	GSFC	3Q 98 3Q 02	
MOD05_L2_G	GSFC	3Q 98 3Q 02	
MOD06_L2_G	GSF	3Q 98 3Q 02	

Files	Archive Site	Epochs Start	End	Model Coverage
	C			
MOD06_L3_MN	GSF C	3Q 98	3Q 02	
MOD07_L2_G	GSF C	3Q 98	3Q 02	
MOD08_L2_G	GSF C	3Q 98	3Q 02	
MOD30_L2_G	GSF C	3Q 98	3Q 02	
MOD38_L2_G	GSF C	3Q 98	3Q 02	
MOD35_L2_G	GSF C	3Q 98	3Q 02	
MOD_ATMOS_L3_ MN	GSF C	3Q 98	3Q 02	
MOD12_L3_3MN	EDC	3Q 98	4Q 98	
MOD12_L3_3MN	EDC	1Q 99	3Q 02	
MOD09_L2_G	EDC	3Q 99	3Q 02	
MOD09_L2_G	EDC	1Q 99	3Q 02	
MOD09_L3_9DY	EDC	3Q 98	3Q 02	
MOD10_L2_G	NSI DC	3Q 98	4Q 98	
MOD10_L2_G	NSI DC	1Q 99	3Q 02	
MOD10_L3_DY	NSI DC	3Q 98	3Q 02	
MOD11_L2_G	EDC	3Q 98	4Q 02	
MOD11_L3_WK	EDC	3Q 98	4Q 02	
MOD12_L3_3MN	EDC	3Q 98	4Q 98	
MOD12_L3_3MN	EDC	1Q 99	4Q 02	
MOD13_L2_G	EDC	1Q 99	3Q 02	
MOD15_L4_WK	EDC	3Q 98	3Q 02	
MOD14_L2_G	GSF C	3Q 98	3Q 02	
MOD14_L3_DY	GSF C	3Q 98	3Q 02	
MOD14_L3_10DY	EDC	3Q 98	3Q 02	
MOD14_L3_MN	EDC	3Q 98	3Q 02	
MOD13_L2_G	GSF C	3Q 98	3Q 02	
MOD16_L3_WK	EDC	3Q 98	3Q 02	

Files	Archive Site	Epochs Start	Epochs End	Model Coverage
MOD17_L4_WK	EDC	3Q 98	3Q 02	
MOD29_L2_G	NSI DC	3Q 98	3Q 02	
MOD29_L3_DY	NSI DC	3Q 98	3Q 02	
MOD33_L3_WK	NSI DC	3Q 98	3Q 02	
MOD34_L3_10DY	EDC	1Q 99	3Q 02	
MOD34_L3_MN	EDC	1Q 99	3Q 02	
MOD40_L3_DY	EDC	1Q 99	3Q 02	
MOD40_L3_10DY	EDC	1Q 99	3Q 02	
MOD40_L3_MN	EDC	1Q 99	3Q 02	
MOD41_L2_MN	EDC	4Q 98	3Q 02	
MOD42_L3_WK	NSI DC	3Q 98	3Q 02	
MODCCLR_L3_CO MP	GSF C	3Q 98	3Q 02	
MODCCLR_L3_CO MP	GSF C	3Q 98	3Q 02	
MOD28_L3_D_DY	GSF C	3Q 98	3Q 02	
MOD28_L3_N_DY	GSF C	3Q 98	4Q 02	
MOD28_L3_D_WK	GSF C	3Q 98	3Q 02	
MOD28_L3_N_WK	GSF C	3Q 98	4Q 02	
MOD32_L2_G	GSF C	3Q 98	3Q 02	

EXHIBIT C-26: MODIS Files

Files	Archive Site	Epochs Start	Epochs End	Model Coverage
MOP- 01	LaRC	2Q 98	4Q 02	
MOP- 02	LaRC	2Q 98	1Q 99	
MOP- 02B	LaRC	2Q 98	1Q 99	
MOP- 02	LaRC	2Q 99	3Q 00	
MOP- 02B	LaRC	2Q 99	3Q 00	
MOP-02	LaRC	4Q 00	3Q 01	
MOP-02B	LaRC	4Q 00	3Q 01	
MOP-03	LaRC	1Q 99	3Q 01	
MOP- 03B	LaRC	1Q 99	3Q 01	

EXHIBIT C-27: MOPITT Files

Files are TBD.

Files	Archive Site	Epochs Start	Epochs End	Model Coverage

EXHIBIT C-28: PR Files

Files	Archive Site	Epochs Start End	Model Coverage
SAG01			
SAG02			
SAG03			
SAG04			
SAG05			
SAG06			
SAG07			
SAG08			
SAG09			
SAG10			

EXHIBIT C-29: SAGE-III Files

Files are TBD.

Files	Archive Site	Epochs Start End	Model Coverage

EXHIBIT C-30: Files

Files	Archive Site	Epochs Start End	Model Coverage
SOL01			
SOL02			
SOL03			
SOL04			

EXHIBIT C-31: SOLTICE Files

Files	Archive Site	Epochs Start End	Model Coverage
SSA01			
SSA02			
SSA03			
SSA04			
SSA05			

EXHIBIT C-32: SSALT Files

Files	Archive Site	Epochs Start End	Model Coverage
SWS01			
SWS02			
SWS03			

EXHIBIT C-33: SWS Files

Files	Archive Site	Epochs Start End	Model Coverage
TES01			
TES02			
TES03			
TES04			
TES05			
TES06			
TES07			
TES08			
TES09			
TES10			
TES11			
TES12			
TES13			
TES14			
TES15			
TES16			
TES17			
TES18			
TES19			
TES20			
TES21			
TES22			

EXHIBIT C-34: TES Files

Files are TBD.

Files	Archive Site	Epochs Start End	Model Coverage

EXHIBIT C-35: TMI Files

Files are TBD.

Files	Archive Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-36: TMR Files

Files are TBD.

Files	Archive Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-37: VIRS Files

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-38: ACRIM Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-39: AIRS Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-40: AMSU Production Processes

Process	Execution Site	Epochs		Model Coverage
		Start	End	
AST_PGE_01	EDC	3Q 98	4Q 02	
AST_PGE_02	EDC	3Q 98	4Q 02	
AST_PGE_03	EDC	3Q 98	4Q 02	
AST_PGE_04	EDC	3Q 98	4Q 02	
AST_PGE_05	EDC	3Q 98	4Q 02	
AST_PGE_06	EDC	3Q 98	4Q 02	
AST_PGE_07	EDC	3Q 98	4Q 02	
AST_PGE_08	EDC	3Q 98	4Q 02	
AST_PGE_09	EDC	3Q 98	4Q 02	
AST_PGE_10	EDC	3Q 98	4Q 02	

EXHIBIT C-41: ASTER Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-42: AVHRR Production Processes

Process	Execution Site	Epochs		Model Coverage
		Start	End	
CER1aT	LaRC	3Q 97	4Q 00	
CER1aA	LaRC	3Q 98	2Q 03	
CER1aP	LaRC	1Q 01	2Q 03	
CER1bA	LaRC	3Q 97	4Q 00	
CER1bP	LaRC	3Q 98	2Q 03	
CER2aT	LaRC	3Q 97	4Q 00	
CER2aA	LaRC	3Q 98	2Q 03	
CER2aP	LaRC	1Q 01	2Q 03	
CER2bT	LaRC	3Q 97	4Q 00	
CER2bA	LaRC	3Q 98	2Q 03	
CER2bP	LaRC	1Q 01	2Q 03	
CER3aA	LaRC	3Q 97	2Q 00	
CER03aA	LaRC	3Q 97	4Q 00	
CER3aP	LaRC	3Q 97	4Q 00	
CER3 bTA	LaRC	3Q 97	4Q 00	
CER3bAP	LaRC	3Q 98	3Q 01	
CER4aV	LaRC	3Q 98	3Q 01	
CER1aF	LaRC	3Q 98	3Q 01	
CER4bA1V	LaRC	1Q 01	3Q 01	
CER4bA1F	LaRC	1Q 01	3Q 01	
CER4bP1V	LaRC	1Q 01	3Q 01	
CER4bP1F	LaRC	3Q 98	3Q 00	
CER5aV	LaRC	3Q 98	3Q 00	
CER5cAV	LaRC	3Q 98	3Q 00	
CER5dPV	LaRC	1Q 01	3Q 01	
CER5aF	LaRC	1Q 01	3Q 01	
CERcAF	LaRC	1Q 01	3Q 01	
CER5dPF	LaRC	3Q 97	4Q 97	
CER5dAF	LaRC	3Q 97	3Q 00	
CER6aT	LaRC	3Q 98	4Q 98	
CER6aA	LaRC	3Q 98	3Q 01	
CER6bA	LaRC	1Q 01	2Q 01	
CER6bP	LaRC	1Q 01	3Q 01	
CER6cT	LaRC	4Q 97	4Q 00	
CER6cA	LaRC	3Q 98	4Q 00	
CER6cP	LaRC	1Q 01	2Q 01	
CER7aA	LaRC	1Q 98	4Q 00	
CER7bA	LaRC	3Q 98	4Q 00	
CER7bP	LaRC	3Q 01	2Q 03	
CER7c	LaRC	1Q 01	2Q 03	
CER7d	LaRC	1Q 98	4Q 00	

Process	Execution Site	Epochs		Model Coverage
		Start	End	
CER8aT	LaRC	3Q 98	2Q 03	
CER8aA	LaRC	1Q 01	2Q 03	
CER8bA	LaRC	1Q 01	2Q 03	
CER8bP	LaRC	1Q 98	4Q 00	
CER8c	LaRC	3Q 98	4Q 00	
CER8d	LaRC	1Q 01	2Q 03	
CER9aTV	LaRC	1Q 01	2Q 03	
CER9aTF	LaRC	3Q 98	4Q 00	
CER9aAV	LaRC	1Q 01	2Q 03	
CER9aAF	LaRC	1Q 98	3Q 00	
CER9aPV	LaRC	1Q 98	4Q 00	
CER9aPF	LaRC	3Q 98	4Q 00	
CER9bTV	LaRC	3Q 98	4Q 00	
CER9bTF	LaRC	3Q 97	4Q 00	
CER9bAV	LaRC	3Q 98	4Q 98	
CER9bAF	LaRC	3Q 98	2Q 03	
CER9bPV	LaRC	1Q 01	2Q 03	
CER9bPF	LaRC	1Q 01	2Q 03	
CER10aT	LaRC	3Q 97	4Q 00	
CER10aA	LaRC	3Q 98	2Q 03	
CER10aP	LaRC	2Q 01	2Q 03	
CER10bTA	LaRC	3Q 98	4Q 00	
CER10bAP	LaRC	1Q 01	2Q 03	
CER11aT	LaRC	3Q 97	4Q 00	
CER11aA	LaRC	3Q 98	2Q 03	
CER11aP	LaRC	1Q 01	2Q 03	
CER12aV	LaRC	2Q 97	3Q 97	
CER12aF	LaRC	3Q 97	2Q 03	

EXHIBIT C-43: CERES Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-44: DORIS Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-45: ESOP Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-46: ETM Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-47: GLAS Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-48: HIRDLS Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	
LIS	MSFC	3Q 97	2Q 03	

EXHIBIT C-49: LIS Production Processes

Processes are TBD.

Process	Execution Site	Epochs Start End	Model Coverage

EXHIBIT C-50: MHS Production Processes

Processes are TBD.

Process	Execution Site	Epochs Start End	Model Coverage

EXHIBIT C-51: MIMR Production Processes

Process	Execution Site	Epochs Start	Epochs End	Model Coverage
MISP1A	LaRC	3Q 98	2Q 03	
MISP1B	LaRC	3Q 98	2Q 03	
MISP1B2	LaRC	3Q 98	3Q 98	
MISP1B2	LaRC	4Q 98	4Q 98	
MISP1B2	LaRC	1Q 99	1Q 99	
MISP1B2	LaRC	2Q 99	2Q 99	
MISP1B2	LaRC	3Q 99	3Q 99	
MISP1B2	LaRC	4Q 99	2Q 03	
MISP2TC	LaRC	3Q 98	3Q 98	
MISP2TC	LaRC	4Q 98	4Q 98	
MISP2TC	LaRC	1Q 99	1Q 99	
MISP2TC	LaRC	2Q 99	2Q 99	
MISP2TC	LaRC	3Q 99	3Q 99	
MISP2TC	LaRC	4Q 99	2Q 03	
MISP2AS	LaRC	3Q 98	3Q 98	
MISP2AS	LaRC	4Q 98	4Q 98	
MISP2AS	LaRC	1Q 99	1Q 99	
MISP2AS	LaRC	2Q 99	2Q 99	
MISP2AS	LaRC	3Q 99	3Q 99	
MISP2AS	LaRC	4Q 99	4Q 99	

EXHIBIT C-52: MISR Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-53: MLS Production Processes

Process	Execution Site	Epochs		Model Coverage
		Start	End	
MOD01:L1A:G	GSFC	3Q 98	3Q 02	
MOD02:L1B:G	GSFC	3Q 98	3Q 02	
MOD03:L1A:G	GSFC	3Q 98	3Q 02	
MOD04:L2:G	GSFC	3Q 98	3Q 02	
MOD04:L3:DY:G	GSFC	3Q 98	4Q 02	
MOD04:L3:WK:G	GSFC	3Q 98	3Q 02	
MOD05:L2:G	GSFC	3Q 98	3Q 02	
MOD06:L3:MN:G	GSFC	3Q 98	3Q 02	
MOD:ATMOS:L2:G	GSFC	3Q 98	3Q 02	
MOD35:L2:G	GSFC	3Q 98	3Q 02	
MOD:ATMOS:L3:MN:G	GSFC	3Q 98	3Q 02	
MOD11:L2:G	GSFC	3Q 98	4Q 98	
MOD11:L2:I	GSFC	1Q 99	3Q 02	
MOD11:L3:WK:G	GSFC	3Q 98	3Q 02	
MOD12:L3:3MN:E	EDC	3Q 98	4Q 98	
MOD12:L3:3MN:I	EDC	1Q 99	3Q 02	
MOD09:L2:G	GSFC	3Q 98	4Q 98	

Process	Execution Site	Epochs Start	Epochs End	Model Coverage
MOD09:L2:I	GSFC	1Q 99	4Q 02	
MOD09:L3:DY:G	GSFC	3Q 98	3Q 02	
MOD10:L2:G	GSFC	3Q 98	4Q 98	
MOD10:L2:I	GSFC	1Q 99	3Q 02	
MOD10:L3:DY:G	NSIDC	3Q 98	3Q 02	
MOD15:L4:WK:G	EDC	3Q 98	2Q 02	
MOD14:L2:G	GSFC	3Q 98	3Q 02	
MOD14:L3:DY:G	EDC	3Q 98	3Q 02	
MOD14:L3:10DY:G	EDC	3Q 98	3Q 02	
MOD14:L3:MN:G	EDC	3Q 98	3Q 02	
MOD13:L2:G	GSFC	3Q 98	3Q 02	
MOD16:L3:WK:G	EDC	3Q 98	3Q 02	
MOD17:L4:WK:G	EDC	3Q 98	3Q 02	
MOD29:L2:G	GSFC	3Q 98	3Q 02	
MOD29:L3:DY:G	NSIDC	3Q 98	3Q 02	
MOD33:L3:WK:G	NSIDC	3Q 98	3Q 02	
MOD34:L3:10DY:I	EDC	1Q 99	3Q 02	
MOD34:L3:MN:I	EDC	1Q 99	3Q 02	
MOD40:L3:DY	EDC	1Q 99	3Q 02	
MOD40:L3:10DY:I	EDC	1Q 99	3Q 02	
MOD40:L3:MN:I	EDC	1Q 99	3Q 02	
MOD41:L2:H	GSFC	4Q 98	3Q 02	
MOD42:L3:WK:G	NSIDC	3Q 98	3Q 02	
MODOCCLR:L2:G	GSFC	3Q 98	3Q 02	
MODOCCLR:SPBIN:G	GSFC	3Q 98	3Q 02	
MODOCCLR:ORBIT:G	GSFC	3Q 98	3Q 02	
MODOCCLR:L3:DY:G	GSFC	3Q 98	3Q 02	
MODOCCLR:L3:WK:G	GSFC	1Q 99	3Q 02	
MODOCCLR:L3:WK:QC:G	GSFC	4Q 98	3Q 02	

Process	Execution Site	Epochs Start	Epochs End	Model Coverage
MOD28:L2:G	GSFC	3Q 98	3Q 02	
MOD28:SPBIN:G	GSF C	3Q 98	3Q 02	
MOD28:D:ORBIT:G	GSF C	3Q 98	4Q 02	
MOD28:N:ORBIT:G	GSF C	3Q 98	4Q 02	
MOD28:L3:COMP:D:D	GSF C	3Q 98	3Q 02	
MOD28:L3:COMP:N:D :Y:G	GSF C	3Q 98	3Q 02	
MOD28:L3:TMP:D:WK	GSF C	3Q 98	3Q 02	
MOD28:L3:TMP:N:WK :G	GSF C	3Q 98	3Q 02	
MOD28:L3:D:WK:G	GSF C	3Q 98	3Q 02	
MOD28:L3:N:WK:G	GSF C	3Q 98	3Q 02	
MOD32:L2:G	GSF C	3Q 98	3Q 02	

EXHIBIT C-54: MODIS Production Processes

Process	Execution Site	Epochs Start	Epochs End	Model Coverage
MOPL1	LaRC	2Q 98	4Q 02	
MOPL1Qi-C	LaRC	2Q 98	3Q 98	
MOPL1Qi-D	LaRC	4Q 98	4Q 02	
MOPL2-C	LaRC	2Q 98	1Q 99	
MOPL2-E	LaRC	2Q 99	2Q 00	
MOPL2-H	LaRC	4Q 00	4Q 02	
MOPL2Qi-C	LaRC	2Q 98	1Q 99	
MOPL2Qi-D	LaRC	2Q 99	3Q 02	
MOPL3	LaRC	1Q 99	3Q 02	
MOPL3Qi-E	LaRC	1Q 99	2Q 99	
MOPL3Qi-F	LaRC	3Q 99	4Q 02	

EXHIBIT C-55: MOPITT Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-56: PR Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-57: SAGE-III Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-58: SeaWiFS Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-59: SOLTICE Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-60: SSALT Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-61: SWS Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-62: TES Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-63: TMI Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-64: TMR Production Processes

Processes are TBD.

Process	Execution Site	Epochs		Model Coverage
		Start	End	

EXHIBIT C-65: VIRS Production Processess

This will account for user workloads in terms of the 49 user services developed from the user scenarios.

User Service	Process(es) Executed	Model Coverage
Simple Search		
Match-up Search		
Coincident Search		
Simultaneous Multiple DAAC Search		
Save Query Results to file for later use		
Spatial Subsetting		
Temporal Subsetting		
Parametric Subsetting		
Spectral Subsetting		
Subset QA statistics		
Save Subsetted data for later "bulk" retrieval		
Save list of lat./long coordinates....		
Access info on non-EOS data		
Access info on EOS and non-EOS algorithms		
Access Electronic Journal		
Integrated Browse (text)		
Cut parts of documents and save to file		
FTP Browse		
Send Browse Products on medium		
Integrated Browse (data)		
Integrated Browse (non-EOS data)		
Display multiple Filess simultaneously		
Animation		
Display "product coverage" map		
Ingest user software or file		
Create and Display 3-D plot		
Create and Display X-Y plot		
Create and Display new images		

User Service	Process(es) Executed	Model Coverage
Create and Display contour plot		
Create and Display scatter plot		
Manage/Save data created by a user process		
Ingest/Archive user results created by processes external to ECS		
Trigger process		
Access Level 0 data		
Access orbital Model output		
Access data dependency info		
Automatic Notification		
Statistical Analysis		
Regrid data		
Coordinate transformation		
Compute difference between two parameters		
Compute ratio of two parameters		
Interactive download		
Point Instrument		
Video Teleconferencing		
Compute order cost		
Order from a saved results list or file:		
Standing Order		
Order results of a user process run on ECS		

EXHIBIT C-66: User/"Pull" Workload

This Exhibit will account for "Push" Workloads by Instrument or other data source; two instruments are listed as examples.

Instrument	"Push" Workload	Process(es) Executed	Model Coverage
ACRIM	Ingest		
	Production		
	Archive		
AIRS	Ingest		
	Production		
	Archive		

EXHIBIT C-67: "Push" Workloads

This Exhibit will account for computer and network resources by subsystem

Subsystem	Processors	Disks	I/O Channels	Network Links	Robots	Read/Write Heads
Client						
Interoperability						
Data Management						
Data Server						
Ingest						
Data Processing						
Planning						

EXHIBIT C-68: System Resources

This Exhibit will account for overheads represented directly in the workloads; an example is given.

Workload	Resource(s)	Model Coverage
Browse	**Data Server Processor at MSFC DAAC**	**DBMS Software Processing**

EXHIBIT C-69: Direct Workload Overhead

This Exhibit will account for background overheads that are not modeled directly. A couple of examples are given.

Resource	Model Coverage
Inter-DAAC Links	Protocol Overhead = 0%
Data Processing Subsystem Processor at GSFC DAAC	Processing Efficiency = 25%

EXHIBIT C-70 : Background Overhead

This Exhibit will account for the data dependencies by process; an example is given.

Process	**Input File(s)**	**Output File(s)**
MOD01:L1A	MOD_L0	MOD01_L1A

EXHIBIT 71: Process Input and Output Files

This Exhibit will account for performance statistics collection for system resources.

Metrics	**Processors**	**Disks**	**I/O Channels**	**Network Links**	**Robots**	**Read/Write Heads**
Number Used						
Utilization Fraction						
Number in Use						
Throughput						
Queue Length						

EXHIBIT C-72: Resource Metrics

This Exhibit will account for performance statistics collection for performance requirements categories.

Performance Requirement	Model Metric
EOSD1000 Emergency real-time commands	
EOSD1010 Appendicies C & D - loads	
EOSD1030 Quick-look data load	
EOSD1040 Reprocessing load	
EOSD1050 Level 1 product availability time	
EOSD1060 Level 2 product availability time	
EOSD1070 Level 3 product availability time	
EOSD1080 Level 4 product availability time	
EOSD1140 Sustaining Engineering resources	
PGS-1300 Processor capacity	
PGS-1310 20% yearly product growth	
PGS-1315 Temporary & Intermediate storage	
PGS-1301 25% of peak CPU capacity	
PGS-1270 PGS expansion by factors of 3,10	
DADS2770 Physical product distribution time	
DADS2778 Receive & archive 3 days' data	
DADS2780 Ingest at max EDOS output	
DADS2900 Archival current + 1 year volume	
DADS2910 Archival storage field expandable	
DADS2950 Manual mounting of archive media	
DADS3000 Bit error rate < 10**(-12)	
DADS3010 Archival/backup media shelf life	
DADS3040 Backup media removable	
DADS3055 Backup media auto/manual mount	
DADS3090 200% throughput expansion	
DADS3100 Network data distribution rates	
DADS3110 Media data distribution rates	
DADS3115 Q-L product distribution time	
DADS3120 QA product distribution time	
DADS3125 Same ECS std-format data	
DADS3126 Different ECS std-format data	
DADS3135 Transaction Rates	
IMS-1780.1 Logon/Authorize response time	
IMS-1780.2 Directory Search response time	
IMS-1780.3 Guide Search response time	
IMS-1780.4 Inventory Search response time	
IMS-1780.5 Status Check response time	
IMS-1780.6 Browse response time	
IMS-1780.7 Document Search response time	
IMS-1780.8 Ordering Services response time	

Performance Requirement	Model Metric
IMS-1785 DADS data base update load	
IMS-1790 Minimum storage - Appendix C	
IMS-1800 Processing/storage expansion	
ESN-1206 Support ECS loads	
ESN-1207 Support ESC growth	

EXHIBIT C-73: Performance Requirements Compliance

APPENDIX D: COST MODEL ANALYSIS DETAIL

D.1 Analysis Methods

In performing the analysis, a common set of evaluation criteria was considered from among the following list:

- Completeness
- Correctness
- Accuracy
- Technical integrity
  - traceability to requirements
  - engineering quality
  - testability
- User satisfaction
  - support for the engineering process
  - implementation

Completeness and correctness, as used in this context, are similar in meaning as when they are used in the context of requirements analysis. Completeness refers to whether all costs are accounted for. Omission of costs leads to unrealistically low estimates and is one source of cost overruns. Correctness, on the other hand, refers to whether the costs included are all required. Addition of costs that are not required obviously inflates the overall estimate, and can result in program termination. However, even if the Cost Modeling is complete and correct, there is no guarantee that it is accurate. Accuracy also requires that the parameters used in estimating the costs be correct.

Technical integrity consists of three components: traceability to requirements; engineering quality; and testability. Traceability to requirements has two facets. First, it refers to whether the implementation of the models satisfies the requirements placed on them by the project. Second, it refers to whether the models fulfill their intended purpose. Engineering quality also has two facets. First, it refers to whether or not good engineering judgment was applied, especially whether the key cost issues and drivers were addressed, and if so, appropriately. The second, and perhaps more important facet of engineering quality, is whether the engineering process uses the models correctly. The third component of technical integrity is testability. Testability in this context, refers to the ease or lack thereof of validating the models.

The last evaluation criteria considered in the analysis was user satisfaction. User satisfaction has two components. The first consideration is whether the model meets its intended purpose. Generally, these models are developed to support the engineering process. If the models do not provide the needed support to the engineering process or are limited in scope, the validity of the answers it provides are in question. The second consideration is related to the first and addresses whether the implementation of the model meets the user's needs and facilitates the purpose it was designed for.

D.1.1 COTS HW and SW Estimation

The specific items addressed in estimating COTS hardware and software costs included:

- Completeness
- Correctness
- Accuracy
  - numbers of components
  - component cost
  - cost trends
- Technical integrity
  - traceability to requirements
  - engineering quality
  - testability
- User satisfaction
  - support for the engineering process
  - implementation

The parameters that tend to drive cost, and are therefore strongly related to accuracy, in the area of COTS hardware and software include numbers of components, the cost per component, and cost trends. Regarding number of components, the key issue addressed was how the numbers were arrived at. This in itself is a complex issue. There was no attempt in this analysis to duplicate the work reported in Section 5. Rather, the goal was to determine whether Performance Modeling was used, and if not, what techniques were used. For component cost, the obvious key issue is whether the costs used for specific components are similar to those available in the marketplace. Given the historical increase in performance and reduction in cost as a function of time, and the fact that any major system, such as EOSDIS is always built over a period of time, the key issue with cost trends is whether that historical performance improvement/cost reduction has been accounted for in estimating costs.

Regarding user satisfaction and the COTS cost estimation, the key issue was whether the model supports its primary purpose; i.e., trade-off analysis. Speed and ease of use are key to achieving that purpose. In contrast, regarding the COTS procurement model, the key issue was its ability to produce accurate costs. For this model, speed and ease of use are less of an issue. Attention must be paid to details regarding the specific components used, their numbers, and their costs.

A variety of methods was used to perform the Cost Model evaluation. The initial plan was to become familiar with the models, their purposes, and the general way they were implemented, and then obtain the models and examine them directly. When it became clear that the models were not going to be made available, the analysis became more focused on learning about the models and performing the evaluation on the basis of that knowledge. Hence, the evaluation was performed by:

- Reviewing existing briefing materials
- Interviewing the model developer
  - conduct interviews

  - write-up findings
  - provide findings to the model developer for comment
  - ask follow-up questions
- Reviewing tabular information provided by the model developer
- Reviewing the PDR modeling plan
- Reviewing documentation of the models produced towards the end of the evaluation process
- Analyzing the implementation of the models based on the developer's description and the other sources of information. Specific issues included
  - inputs
  - decision criteria
  - outputs
- Estimating the historical rate of decrease of cost per unit of performance

In the process of the evaluation some specific "measurements" were made. The types of values examined included:

- parameters (e.g. rates of decrease in performance per unit price)
- the specific hardware / software selections used as the basis of cost

The historical rate of decrease in the cost per unit of performance was estimated by consulting a variety of references such as old purchase orders and old magazine articles for historical cost data, current list prices and vendor quotes for current costs, and data bases of benchmark data for performance information. Two to three specific examples were utilized for each technology area. The percent decrease in performance per unit price per year was computed for each example. These values were then compared with those being used by HAIS. A sensitivity analysis was also performed to determine what parameters would be required to be input to achieve the percent decrease in performance per unit price per year being used by HAIS. These parameters were then evaluated to determine whether they were reasonably close to known price/performance points or not.

The formula for computing cost/capacity decrease per year is:

Cost/Capacity decrease per year =	Initial Cost / Capacity minus Cost / Capacity in Year 1 divided by the Initial Cost / Capacity
Cost/Capacity in Year 1 =	Initial Cost / Capacity times the annual Cost / Capacity reduction factor
Annual Cost/Capacity reduction factor =	The Nth root of the ratio of Initial Cost / Capacity to Ending Cost Capacity where N equals the number of years between the two data points

Data points input to these formula as well as the results of the computations are shown.

	Start Value	End Value	Year 1 Value	Year 2 Value	Year 3 Value	Year 4 Value	Year 5 Value	Number of Years (N)	Nth Root
CPU - estimated values									
Capacity (MIPS)	50	200							
Cost	20000	10000							
Cost/Capacity	400.00	50.00	263.90	174.11	114.87	75.79	50.00	5	65.98
Percent Decrease			-0.34	-0.34	-0.34	-0.34	-0.34		
CPU - another estimate									
Capacity (MIPS)	4	40							
Cost	2500	2000							
Cost/Capacity	625.00	50.00	377.14	227.57	137.32	82.86	50.00	5	60.34
Percent Decrease			-0.40	-0.40	-0.40	-0.40	-0.40		
CPU - Based on published numbers									
Capacity (Mflops)	6	18							
Cost	28000	24000							
Cost/Capacity	4666.67	1333.33	3073.62	2024.39	1333.33			3	65.86
Percent Decrease			-0.34	-0.34	-0.34				
DISK - based on cost estimates									
Capacity (MB)	1000	4000							
Cost	7000	3200							
Cost/Capacity	7.00	0.80	4.54	2.94	1.90	1.23	0.80	5	64.80
Percent Decrease			-0.35	-0.35	-0.35	-0.35	-0.35		
DISK - Based on Actual Purchases									
Capacity (MB)	0.1	0.4							
Cost	575	220							
Cost/Capacity	5750.00	550.00	3197.73	1778.34	988.98	550.00		4	55.61
Percent Decrease			-0.44	-0.44	-0.44	-0.44			
ARCHIVE - based on cost estimates									
Capacity (TB)	1	6							
Cost	200	150							
Cost/Capacity	200.00	25.00	131.95	87.06	57.43	37.89	25.00	5	65.98
Percent Decrease			-0.34	-0.34	-0.34	-0.34	-0.34		
ARCHIVE - based on vendor quotes									
Capacity (TB)	1	14.5							
Cost	200	625							
Cost/Capacity	200.00	43.10	136.27	92.85	63.26	43.10		4	68.14
Percent Decrease			-0.32	-0.32	-0.32	-0.32			

EXHIBIT D-1: Data Points Used To Compute Annual Price / Performance Decrease And The Resulting Values

D.1.2 Custom SW Estimation

The custom software estimation analysis focused on an assessment of the following:

- Accuracy
  - numbers of elements
  - size multipliers
  - cost trends
- Technical integrity
  - traceability to requirements
  - engineering quality
  - testability
- User satisfaction
  - support for the engineering process
  - implementation

Completeness and correctness could not be evaluated due to a lack of information. The majority of the effort was focused on assessing the engineering quality of the estimation process. Software cost estimation follows four basic steps:

- size estimation
- level of effort estimation
- schedule determination
- cost estimation

Due to lack of available information, the analysis focused on determining how the steps mentioned above were performed. It was determined early in the evaluation process that a somewhat new method was being used to estimate software size. Therefore, a key aspect of the analysis addressed the validity of that method. Some of the specific tasks performed included:

- Determining what methods exist for estimation of software size/effort given an object-oriented design process and the degree to which they have been tested;
- Searching for an established relationship between source lines of code (SLOC) and object-oriented design (OOD) entities; and
- Determining the relationship between SLOC and OOD entities such as objects and methods in existing C++ software.

As discussed in Section 6.1, even if a valid method is used, the parameters used within the method also strongly effect whether the costs estimated are accurate or not. Hence, this analysis examined three key parameters: numbers of elements; size multipliers; and cost trends to determine whether they seemed to be valid and/or were being estimated in a valid way. Numbers of elements in this context refer to counts of object-oriented design entities, such as objects (similar to structures in C) and methods (similar to functions in C). Size multipliers refer to the number of SLOC used to multiply against the counts of design entities. Cost trends in this context refer to whether changes in cost (personnel cost in this case) were recognized and accounted for correctly.

The other issues included in evaluation of technical integrity were similar to those discussed in Section 6.1.

The analysis was performed primarily by interviewing the model developer and analyzing the stated approach, as performed for COTS hardware and software estimation above. Due to difficulties in coordination with this individual, it was not possible to conduct several iterations of this process. Moreover, no hardcopy materials were provided to aid in the analysis. Since, the estimation approach being used is somewhat new, the analysis did include querying the literature and engineering community for existence/familiarity with stated the approach and for recommended strategies.

In the process of the evaluation some specific "measurements" were made. The types of values that were examined were:

- parameters (e.g. SLOC / OOD entity)

Two methods were used to determine whether methods exist for estimation of software size/effort given an object-oriented design process and what relationships have been established between source lines of code (SLOC) and object-oriented design (OOD) entities. First the Internet news system was used to circulate questions to the software engineering community. A synopsis of the EOSDIS project, the phase of the development and the stated approach was placed in the news system with a request for comments as to knowledge of projects in which the stated approach had been used and whether any tools existed that had successfully parameterized the approach. Second, to insure that the existing knowledge base was being thoroughly sampled, Peter Coad, an author of several books on object-oriented design, was contacted for existing materials on the subject of object-oriented estimation methods. The readily available references returned by these two methods were obtained and examined for relevant information. In addition, written guidelines for software estimation published by Reifer Consultants were used.

In order to determine the relationship between SLOC and OOD entities such as objects and methods in existing software, two repositories of C++ software were analyzed. Ratios were computed by counting SLOC within each class with a code counter, counting methods within each class manually, then computing the ratios.

D.1.3 Operational Cost Estimation

The specific issues addressed in analysis of estimation of operational costs included:

- Completeness
- Correctness
- Accuracy
    - rates
- Technical integrity
    - traceability to requirements
    - engineering quality

  - testability
- User satisfaction
  - support for the engineering process
  - implementation

Operational costs are driven by personnel costs. Therefore, the key issues regarding estimation of operational costs center around how personnel costs are estimated. Completeness in this context refers to whether all the required types of staff were accounted for. Correctness refers to whether the types of personnel included are all required.

The engineering quality aspects of the technical integrity assessment had some unique aspects. The key issues related to how the numbers of personnel were estimated and whether the schedule for delivery of automated features was factored into the estimation process.

The analysis was performed primarily by interviewing the model developer and analyzing the stated approach, as reported for COTS hardware and software estimation above. It was only possible to conduct two iterations of this process. The staff allocations performed for SDR were provided to aid in understanding the organization of staff functions. However, the numbers of staff reflected in this material were deemed to be outdated and invalid, and therefore, were not analyzed.

Early in the analysis it was determined that a fixed value of $100K was being assumed for the cost of a man-year. In order to evaluate the validity of using such a figure, a typical mix of labor categories and costs per labor category was assumed. The average salary resulting from these assumptions was then computed and compared to the stated value of $100K. The inputs to this analysis of average annual labor cost are shown in Exhibit D-2.

POSITION TITLE	Number of Staff	Cost/HR (unloaded)	Multiplier used	ANNUAL COST	TOTAL COST
M&O Manager	4	$40	off-site	$188,416	$753,664
Site Manager	1	$35	on-site	$143,360	$143,360
Admin Support / Security	3.7	$15	on-site	$61,440	$227,328
Librarian	1	$15	on-site	$61,440	$61,440
Operational Readiness and Performance Assurance	2	$20	on-site	$81,920	$163,840
DAAC Trainers	2	$20	on-site	$81,920	$163,840
SUSTAINING ENGINEERING					
S/W Maintenance/Engineering	20.2	$30	off-site	$141,312	$2,854,502
Science	2	$25	on-site	$102,400	$204,800
Planned Upgrades	1.7	$25	on-site	$102,400	$174,080
Configuration Management	7.2	$25	on-site	$102,400	$737,280
Testing	3	$25	on-site	$102,400	$307,200
Property Management/ILS	4.4	$15	on-site	$61,440	$270,336
H/W Maintenance	10	$30	on-site	$122,880	$1,228,800
Resource Controller/Performance Analyst	1	$25	on-site	$102,400	$102,400
ALGORITHM SUPPORT					
Test & Integration	6	$20	on-site	$81,920	$491,520
Development	4	$25	off-site	$117,760	$471,040
Data Base Administration	2	$20	on-site	$81,920	$163,840
Ops. Supervisor/Production Scheduler	1	$20	on-site	$81,920	$81,920
QA/Production Monitor	22.1	$20	on-site	$81,920	$1,810,432
Ground Controller	7.2	$20	on-site	$81,920	$589,824
USER SERVICES					
Data Specialist	10.2	$20	on-site	$81,920	$835,584
User Assistance	13	$15	on-site	$61,440	$798,720
Data Distribution Technician	8.2	$15	on-site	$61,440	$503,808
Computer Operator	10.3	$15	on-site	$61,440	$632,832
Archive Manager	4.7	$20	on-site	$81,920	$385,024
TOTALS	151.9				$14,157,414
		=	**Average cost per man-year**		**$93,202**
onsite_indirect_rate	2				
offsite_indirect_rate	2.3				

EXHIBIT D-2: Parameters Used In Estimating The Average Cost Per Man Year

D.2 Analysis Results

The results of analysis of traceability to requirements is shown in Exhibit D-3.

REQUIREMENT	SATISFIED?	COMMENTS
The contractor shall establish and maintain a Life Cycle Cost (LCC) model	PARTIAL	There is no overall model which computes lifecycle cost. Rather, outputs from a series of models must be manually accumulated.
The LCC model shall be developed to be compatible with the ECS Work Breakdown structure (WBS).	TBD	
The LCC model shall identify lifecycle costs including the cost of development, acquisition, operation, COTS licensing, upgrades (including newer versions of COTS software), correction of latent defects, and related system support over the ECS lifetime.	PARTIAL	The components of cost described in the requirement are being estimated individually. There is, however, no overall model which rolls up these components of cost into an overall lifecycle cost.
The LCC model shall also include the cost of any necessary maintenance subcontracts.	YES	
The LCC model shall include projections for technology improvements.	YES	
The contractor shall provide ECS Life cycle Cost Reports in accordance with DID 213/SE2.	TBD	
The LCC model shall model cost sensitive parameters to provide the Government with the capability assess cost and schedule impacts of new or modified requirements. Cost sensitive parameters shall include, but not be limited to: new instruments, schedule changes, processing requirements, archive volume requirements, number of granules, number of products, and input/output loads.	PARTIAL	View #1: Since there is no standalone model, there is no capability for the government to do this. View #2: The "Interactive Cost Model" partially fulfills this requirement. The write-up in the "PDR Modeling Plan" falls short of the SOW requirements.
The LCC Model shall be continuously updated with actual performance data.	YES	Changes have been made between SDR and PDR.
The LCC Model, as well as the results from it, shall be made available to the Government.	PARTIAL	Results and some parts of model have been delivered. However, the entire model, in fact, cannot be made available, since it does not exist in standalone form.

EXHIBIT D-3: Cost Modeling Requirements Satisfaction Matrix

APPENDIX E: LIST OF REFERENCES

The following documents were referenced in the assessment of the ECS models or in the preparation of this report:

1. *EOSDIS Modeling Assessment Plan,* Intermetrics, EOSDIS IV&V , Intermetrics, September 14, 1994
2. *EOSDIS Modeling Assessment Report (Draft Preliminary),* EOSDIS IV&V deliverable #IVV-0506, Intermetrics, October 7, 1994
3. *Functional and Performance Requirements Specifications for the ECS,* Revision A, #423-41-02, HAIS, June 2, 1994
4. *ECS Design Specifications for the ECS Project (Draft, Preliminary),* #194-207-01, HAIS, July 1994
5. *ECS User Scenario Notebook,* Technical Paper #194-00311TPW, HAIS, June 1994
6. *ECS User Characterization Methodology Results,* White Paper #194-00313TPW, HAIS, September 1994
7. *User Scenario Functional Analysis,* White Paper #194-00548TPW, HAIS, October 1994
8. *User Characterization and Requirements Analysis,* White Paper #194-00312TPW, HAIS, September 1994
9. AHWGP Workshop Presentation, at HAIS, September 1994
10. AHWGP Workshop Notes, A. Sanyal, SMSRC/Intermetrics, December 1994
11. AHWGP input files available from the EDHS/AHWGP ftp server, 1994–1995
12. *Science Operations Concepts for EOSDIS: Part 1, Data Products Resource Allocation, version 1.0*, (Draft Plan), S. Wharton and M. Myers, GSFC, August 1994
13. *EOSDIS Capacity Allocation,* IWG Meeting Presentation by S. Wharton, GSFC, October 1994
14. *EOSDIS Output Data Products and Input Requirements; Interim Version,* SPSO/GSFC, July 1994
15. *EOSDIS Output Data Products an Input Requirements, version 2.0, Volume II: Analysis of IDS Input Requirements,* SPSO/GSFC, August 1992
16. *Science Data Plan for the EOSDIS (draft),* Matthew Schwaller and Brian Krupp, GSFC, August 1994
17. *Algorithm Theoretical Basis Documents (ATBD),* EOS Instrument Scientists, compiled by SPSO/GSFC, January 1995
18. *SDPS Performance Requirements Interpretation,* HAIS, June 1994
19. *HAIS PDR Modeling Plan, Version 1.1,* (draft) Mark Settle, HAIS, October 1994
20. HAIS Monthly Status Meeting Information Packages, August–December, HAIS, 1994
21. *ECS Scientist User Survey (ESUS),* Technical paper #194-00549TPW, HAIS, October 1994

22. *ECS User Characterization and Results,* HAIS, May 1994
23. *BONeS Designer User's Guide (for use with version 2.6),* COMDISCO Systems, December 1993
24. *BONeS Designer Modeling Reference Manual (for use with version 2.6),* COMDISCO Systems, December 1993
25. *BONeS Designer Core Library Guide (for use with version 2.6),* COMDISCO Systems, December 1993
26. *COTS Cost Estimation Model for the ECS Project,* Technical Paper # 231-TP-001-001, HAIS, December 1994
27. *Bill of Materials Procurement Cost Model for the ECS Project,* Technical Paper #231-TP-002-001, HAIS, December 1994
28. *A Preliminary EOSDIS User Model*, Bruce R. Barkstrom, unpublished manuscript, NASA Langley Research Center, Hampton, VA, 1991
29. System Design Review (SDR), presented at GSFC on June 27-28, 1994
30. *Software Resource Estimating Procedure,* Donald J. Reifer, Reifer Consultants, Inc., March 1992
31. *Leveraging Object-Oriented Development at AMES*, Greg Wenneson and John Connell, SEPG, Sterling Software at NASA Ames,

APPENDIX F: TOOLS AND DATA BASES UTILIZED

This section describes the data bases and tools used by IV&V team to assess the User, Production, Performance and Cost Models. Specific data bases and tools used, including specific version and operational environment, are provided in Exhibit F-1.

IV&V TOOLS	ENVIRON-MENT	SPECIFIC DATA BASES	MODEL
MS Excel 5.0	PCs	HAIS EOSDIS User Scenarios AHWGP Processing Time lines v2.0 AHWGP Processing Descriptions v2.0 AHWGP Volume Timelines v2.0 AHWGP File Descriptions v2.0 ATBD Data Product Summary SPSO Science Data Plan SPSO Output Data Products & Input Req.	User Production Production Production Production Production Production Production
MS Word	PCs		All
MS Access	PCs		All
MS PowerPoint	PCc		All
Lotus Approach 3.0	PCs	MTPE Data Base MTPE Landsat User Data Base MTPE USGCRP Research Fellowship Data Base	User User User
BONeS 2.6.1	SUN, VT 110, 220 (Emulator)		Performance Production
Novell Netware LAN/WAN LAN WorkPlace Internet	 PCs SUN, PCs, and Mac SUN, PCs SUN, PCs, Mac		 All

EXHIBIT F-1: Tools And Data Bases Used

In addition, the IV&V Cost Model analyses utilized several specialized external tools and data bases:

COTS H/W and S/W Estimation - External sources were consulted to determine the realistic trends in price versus performance. For CPU performance vs. time, Business Week, July 4, 1994 and NETLIB at Oak Ridge National Labs were consulted. For archive capacity vs. time the National Media Labs Independent Report was utilized.

Custom S/W Estimation - The comp.software-eng news group within the Internet news system was used to query the software engineering community regarding estimation methods for use with object oriented design.

Two software repositories were used to determine the typical values for numbers of SLOC per method: They are shown in Exhibit F-2.

REPOSITORY	ORGANIZATION	DESCRIPTION
NIH Class Library	National Institutes of Health	The library described in the book "Data Abstraction and Object-Oriented Programming in C++" by Keith E. Gorlen, Sanford M. Orlow, and Perry S. Plexico published by John Wiley and Sons.
FAST PPS (based on the Telemetry Processing Control Environment (TPCE))	GSFC, Software and Automation Systems Branch (Code 522)	Code implementing the packet processing system for the Far Auroral Snapshot Explorer mission
SAMPEX CMS	GSFC, Software and Automation Systems Branch (Code 522)	Code implementing the command management system for the Solar Anomalous and Magnetosphere Particle Explorer mission

EXHIBIT F-2: Software Repositories Used To Determine SLOC

A CTA proprietary code counting utility named "sloc" was used to count source lines of code within the c++ repositories. This utility has been tested against manual counts and found to be extremely accurate.

Operational Cost Estimation - The comp.software-eng news group within the Internet news system was also used to determine the availability of tools for estimation of O&M costs.